



UNIVERSITAT POLITÈCNICA DE CATALUNYA

TRABAJO DE FINAL DE GRADO

Análisis estadístico de diferentes estrategias de apuestas: Aplicación a datos de fútbol.

Autor:

David SEDEÑO CASTAÑOS

Directora:

Dr. Lesly María ACOSTA

Codirector:

Dr. Josep Antón SÁNCHEZ



Grau en Enginyeria en Tecnologies Industrials

Escola Tècnica Superior d'Enginyeria Industrial de Barcelona (ETSEIB)

Departamento de Estadística e Investigación Operativa

Setiembre 2018

Agradecimientos a mis directores

Dra. Acosta Argueta & Dr. Sàncchez Espigares

por las aportaciones técnicas.

A familia y amigos

por la presencia, el apoyo y la confianza.

Resumen

Análisis estadístico de diferentes estrategias de apuestas: Aplicación a datos de fútbol

por David SEDEÑO CASTAÑOS

A modo de finalizar el grado, llegaba el momento de escoger un tema para el trabajo de fin de grado, donde demostrar algunos de los conocimientos adquiridos, así como también la capacidad de búsqueda y resolución. Tras pensar en varios de los campos que se han tocado durante estos cuatro años, se acabó optando por realizar algún análisis de datos, y el tema vino después.

El objetivo del proyecto, a parte del desarrollo personal, era encontrar alguna estrategia con la que incrementar los beneficios realizando apuestas de fútbol en la liga española. Más concretamente, se pretende poder ayudar a alguien sin conocimientos de fútbol a poder obtener mejores resultados en términos de beneficios, que si apostase al azar. Se hace hincapié durante el trabajo en que no es transcendental aumentar el porcentaje de aciertos, pues al tener cada partido un beneficio diferente, ello no implicaría mejores resultados en términos económicos.

Para poder conocer si había mejoras o no, era necesario establecer una referencia, y que aquí se las nombró “estrategias básicas”. Es por ello, que se comenzó con análisis y descripción del tema y se siguió con la aplicación de las mencionadas estrategias básicas. Eran necesarios estos pasos tanto para poder entender el proyecto como para comparar los resultados de los diseños con los iniciales. Para realizar un estudio riguroso, las estrategias se extraen a partir de los datos de la temporada 2016-17, y se prueban sobre los datos de la 2017-18.

En un inicio se confió en poder establecer alguna estrategia que mejorara los resultados de las básicas, pero se pensaba que aun así no llegaría a alcanzar beneficios económicos. Se comenzaron probando estrategias ya conocidas, seguido de estrategias extraídas de la exploración de datos, y se finalizó con la obtención de varias estrategias a partir del uso de herramientas y técnicas estadísticas sofisticadas. Todas ellas se volvieron a probar con datos diferentes.

Se ha acabado logrando mejoras respecto a las situaciones básicas, e incluso considerables beneficios.

La elección de este tema se ha considerado como un acierto, ha aportado importantes conocimientos, además de descubrir y manejar herramientas de gran utilidad.

Índice general

| | |
|--|-------------|
| Resumen | V |
| Índice general | VII |
| Índice de figuras | IX |
| Índice de cuadros | XIII |
| Lista de abreviaciones | XV |
| Glosario | XVII |
| 1. Prefacio | 1 |
| 1.1. Origen del proyecto | 1 |
| 1.2. Motivación | 1 |
| 2. Introducción | 3 |
| 2.1. Objetivos del proyecto | 3 |
| 2.2. Alcance del proyecto | 3 |
| 3. Apuestas y cuotas | 5 |
| 3.1. Breve introducción a las apuestas | 5 |
| 3.1.1. Preguntas aún sin respuesta | 5 |
| 3.2. Tema a tratar | 6 |
| 3.3. Fuente de datos | 6 |
| 3.4. Contenido del fichero | 7 |
| 3.5. Funcionamiento de las cuotas | 9 |
| 4. Primeros análisis | 11 |
| 4.1. Resultado final | 11 |
| 4.2. Favorito | 13 |
| 4.3. Transformación del valor de cuota a probabilidad de éxito | 17 |
| 5. Estrategias | 19 |
| 5.1. Estrategias simples | 19 |
| 5.2. Estrategias existentes | 21 |

| | |
|---|-----------|
| 5.2.1. Surebet | 21 |
| 5.2.2. Martingala | 31 |
| Martingala en el fútbol | 32 |
| 5.3. Estrategias con exploración de datos: Caracterización de las casas de apuestas | 37 |
| 5.3.1. Empate | 40 |
| 6. Diseño de estrategias a partir de técnicas de análisis de datos | 49 |
| 6.1. Análisis de componentes principales | 49 |
| 6.2. Cluster k-Means | 52 |
| 6.2.1. Resultados | 54 |
| 6.3. Análisis discriminante | 63 |
| 6.3.1. Resultados | 64 |
| Separación lineal | 64 |
| Separación cuadrática | 66 |
| 6.4. Regresión Logística Nominal | 71 |
| 6.4.1. Resultados | 73 |
| Probabilidad corregida | 76 |
| 7. Planificación | 79 |
| 8. Presupuesto | 81 |
| 9. Conclusiones | 83 |
| 10. Impacto medioambiental | 87 |
| 11. Agradecimientos | 89 |
| Bibliografía | 91 |
| A. Base de datos | 93 |
| B. Planificación | 97 |
| B.1. Calendario teórico | 98 |
| B.2. Calendario práctico | 102 |

Índice de figuras

| | |
|---|----|
| 3.1. Captura de las cuotas tipo 1X2 de las dos primeras jornadas | 9 |
| 4.1. Diagrama de Pareto de los resultados de final de partido | 11 |
| 4.2. Diagrama de Pareto de los resultados en la media parte | 12 |
| 4.3. Diagrama de Pareto de los resultados finales estratificados según resultado al descanso | 12 |
| 4.4. Captura de las cuotas tipo 1X2 y de quién es favorito. | 13 |
| 4.5. Diagrama de Pareto de qué equipo fue favorito | 14 |
| 4.6. Diagrama de Pareto de resultado final cuando el favorito es el local | 14 |
| 4.7. Diagrama de Pareto de resultado final cuando el favorito es el visitante | 15 |
| 4.8. Captura de Minitab de la prueba de hipótesis para apuestas al favorito | 15 |
| 4.9. Dotplot de las cuotas del Real Madrid cuando fue local y además ganó | 16 |
| 5.1. Captura de tabla con las rentabilidades ofrecidas por Bet365 | 24 |
| 5.2. Histograma de las rentabilidades en Bet365 | 24 |
| 5.3. Captura de tablas con rentabilidades en las 7 casas de apuestas | 25 |
| 5.4. Tabla con la mejor cuota de cada tipo, y con la nueva rentabilidad | 26 |
| 5.5. Dotplot del beneficio neto (en tanto por 1) en presencia de surebet favorable | 26 |
| 5.6. Estadísticas descriptivas de la ventaja con surebet | 26 |
| 5.7. Captura de tablas de casas de apuestas con mejores cuotas | 27 |
| 5.8. Casas de apuestas que mejor pagaban al local. | 28 |
| 5.9. Casas de apuestas que mejor pagaban el empate. | 28 |
| 5.10. Casas de apuestas que mejor pagaban al visitante. | 29 |
| 5.11. Casas de apuestas que ofrecían la mejor cuota del resultado que finalmente se dió | 29 |
| 5.12. Captura de tabla de rentabilidad sin la casa de apuestas Pinnacle. | 30 |
| 5.13. Dotplot de la ventaja de la surebet sin Pinnacle. | 30 |
| 5.14. Estadísticas descriptivas de la ventaja sin Pinnacle | 31 |
| 5.15. Estadísticas descriptivas de las máximas cuotas para cada tipo de apuesta | 33 |
| 5.16. Diagrama de Pareto del resultado final, para seleccionar qué martingala utilizar | 33 |
| 5.17. Estadísticas descriptivas de la cuota pagada en los tipos 1, X y 2 | 34 |
| 5.18. Captura de tabla de las cuentas para una martingala al local | 35 |
| 5.19. Captura de tabla de las cuentas para una martingala al empate | 36 |

| | |
|--|----|
| 5.20. Captura de tabla de las cuentas para una martingala al local al visitante . . . | 37 |
| 5.21. Dispersión de datos en plano PC2-PC1 y ejes de las variables | 39 |
| 5.22. Diagrama con partidos pinados según FTR en el plano PC2-PC1 | 39 |
| 5.23. Histograma del valor de las cuotas del empate. | 41 |
| 5.24. Histograma del valor de las cuotas del empate, diferenciando según si se dio empate o no | 41 |
| 5.25. Histograma de los valores de las cuotas, con frecuencias y con separación según acierto | 42 |
| 5.26. Captura de la operativa. Importancia de parametrizar. | 43 |
| 5.27. Evolución de los beneficios de la estrategia de caracterización conseguida, aplicación sobre las últimas 4 temporadas | 47 |
| | |
| 6.1. Ejemplo de las dos direcciones principales de inercia para una dispersión de puntos. | 51 |
| 6.2. Análisis de componentes principales | 52 |
| 6.3. Dispersión de las probabilidades en el plano PC2-PC1 | 52 |
| 6.4. Muestra de centroides de grupos | 53 |
| 6.5. Iteraciones en búsqueda de convergir. | 54 |
| 6.6. Grupos creados pintados. Plano PC1-PC2. | 55 |
| 6.7. Captura de la tabla de correspondencia que relaciona grupos y resultados . . | 55 |
| 6.8. Evolución de beneficios acumulados apostando únicamente al grupo 1 | 56 |
| 6.9. Evolución de beneficios acumulados apostando únicamente al grupo 2 | 56 |
| 6.10. Evolución de beneficios acumulados apostando únicamente al grupo 3 | 57 |
| 6.11. Evolución de beneficios acumulados apostando únicamente al grupo 4 | 57 |
| 6.12. Información de los grupos creados. Observaciones, centroides y distancias entre grupos. | 59 |
| 6.13. Evolución de beneficios acumulados apostando únicamente al grupo 1. Fase de pruebas. | 61 |
| 6.14. Evolución de beneficios acumulados apostando únicamente al grupo 2. Fase de pruebas. | 61 |
| 6.15. Evolución de beneficios acumulados apostando únicamente al grupo 3. Fase de pruebas. | 62 |
| 6.16. Evolución de beneficios acumulados apostando únicamente al grupo 4. Fase de pruebas. | 62 |
| 6.17. Captura de la tabla de correspondencia de los grupos creados y de los grupos reales. Separación lineal | 65 |
| 6.18. Captura de la tabla de correspondencia de los grupos creados y de los grupos reales. Separación lineal. Validación cruzada. | 65 |
| 6.19. Evolución de los beneficios acumulados con validación cruzada y separación lineal | 66 |

| | |
|---|-----|
| 6.20. Captura de la tabla de correspondencia de los grupos creados y de los grupos reales. Separación cuadrática. | 67 |
| 6.21. Captura de la tabla de correspondencia de los grupos creados y de los grupos reales. Separación cuadrática. Validación cruzada. | 67 |
| 6.22. Evolución de los beneficios acumulados con validación cruzada y separación cuadrática | 67 |
| 6.23. Evolución de los beneficios acumulados apostando al local con el modelo extraído del análisis discriminante. Fase de pruebas. | 69 |
| 6.24. Evolución de los beneficios acumulados apostando al empate con el modelo extraído del análisis discriminante. Fase de pruebas. | 70 |
| 6.25. Evolución de los beneficios acumulados apostando al visitante con el modelo extraído del análisis discriminante. Fase de pruebas. | 70 |
| 6.26. Tabla de regresión logística con todas las variables | 73 |
| 6.27. Captura de tabla de regresión logístca con las variables significativas | 74 |
| 6.28. Evolución de los beneficios acumulados aplicando la estrategia obtenida con NLR | 74 |
| 6.29. Evolución de los beneficios acumulados aplicando la estrategia obtenida con NLR. Fase de pruebas. | 76 |
| 6.30. Evolución de los beneficios acumulados aplicando la estrategia obtenida con NLR | 77 |
| 6.31. Evolución de los beneficios acumulados aplicando la estrategia obtenida con NLR. Fase de pruebas. | 78 |
| 9.1. Evolución de las estrategias básicas | 85 |
| 9.2. Evolución de las estrategias diseñadas | 85 |
| A.1. Base de datos utilizada. Parte 1. | 94 |
| A.2. Base de datos utilizada. Parte 2. | 95 |
| A.3. Base de datos utilizada. Parte 3. | 96 |
| B.1. Diagrama de Gantt teórico. Parte 1. | 98 |
| B.2. Diagrama de Gantt teórico. Parte 2. | 99 |
| B.3. Diagrama de Gantt teórico. Parte 3. | 100 |
| B.4. Diagrama de Gantt práctico. Parte 1. | 102 |
| B.5. Diagrama de Gantt práctico. Parte 2. | 103 |
| B.6. Diagrama de Gantt práctico. Parte 3. | 104 |
| B.7. Diagrama de Gantt práctico. Parte 4. | 105 |

Índice de cuadros

| | |
|---|----|
| 3.1. Ejemplo cuotas | 9 |
| 3.2. Ejemplo de unas aplicaciones de cuotas para casos hipotéticos | 10 |
| 4.1. Ejemplo de la consideración de favorito | 13 |
| 4.2. Resultado de apostar toda una temporada al favorito | 16 |
| 4.3. Resultado de apostar a un favorito toda la temporada | 16 |
| 4.4. Ejemplo para transformar el valor de cuota a probabilidad de éxito | 17 |
| 4.5. Cuotas y probabilidades. | 17 |
| 4.6. Probabilidades de éxito corregidas | 18 |
| 5.1. Apuesta simple al local (L). | 19 |
| 5.2. Apuesta simple al visitante (V). | 20 |
| 5.3. Apuesta simple al empate (E). | 20 |
| 5.4. Apuesta simple al local cuando es favorito (L1). | 20 |
| 5.5. Apuesta simple al visitante cuando es favorito (V1). | 20 |
| 5.6. Resumen de los resultados de cada estrategia simples | 21 |
| 5.7. Ejemplo de como realizar una Surebet | 22 |
| 5.8. Aplicación de la condición de beneficios iguales en Surebet | 22 |
| 5.9. Resultado del ejemplo de Surebet propuesto | 22 |
| 5.10. Ejemplo de aplicación de Martingala | 31 |
| 5.11. Ejemplo de cuotas. Importancia de la diferencia de cuotas de local y visitante | 42 |
| 5.12. Resultados para la primera parametrización | 44 |
| 5.13. Resultados de varias estrategias con mejores resultados sobre la propia temporada | 45 |
| 5.14. Resultados de la estrategia de caracterización escogida, aplicación en las últimas cuatro temporadas | 46 |
| 6.1. Resumen de las técnicas que se emplearán | 49 |
| 6.2. Beneficios obtenidos según grupos y apuesta a realizar, para k-means | 58 |
| 6.3. Resultados de apuestas interesante con k-means. | 58 |
| 6.4. Apuestas aconsejadas, extraídas de k-means | 58 |
| 6.5. Resultados de la fase de pruebas con k-means | 60 |
| 6.6. Beneficio de las apuestas a los 3 resultados con técnica de análisis discriminante y separación lineal | 66 |

| | |
|---|----|
| 6.7. Beneficio de las apuestas a los 3 resultados con técnica de análisis discriminante y separación cuadrática | 68 |
| 6.8. Beneficio de las apuestas al local con técnica de análisis discriminante y separación cuadrática | 68 |
| 6.9. Resumen de las apuestas realizadas y aconsejadas con uso de técnicas de análisis discriminante. Fase de pruebas. | 71 |
| 6.10. Resumen de la estrategia implementada con NLR | 75 |
| 6.11. Resumen de la estrategia implementada con NLR. Fase de pruebas. | 76 |
| 6.12. Resumen de la estrategia implementada con NLR con corrección | 77 |
| 6.13. Resumen de la estrategia implementada con NLR con corrección. Fase de pruebas. | 77 |
| 8.1. Recursos utilizados y costes asociados | 81 |
| 8.2. Resumen de costes estimados | 82 |
| 9.1. Resumen de las mejoras conseguidas en forma de esperanza matemática | 84 |

Lista de abreviaciones

| | | |
|----------------|--|---|
| H | Home Team/Win | Local (Equipo/Victoria) |
| D | Draw | Empate |
| A | Away Team/Win | Visitante (Equipo/Victoria) |
| PCA | Principal Components Analysis | Análisis de components principales |
| PC | Principal Components | Componentes principales |
| NLR | Nominal Logistic Regression | Análisis de regresión logística |
| QDA | Quadratic Discriminant Analysis | Análisis discriminante cuadrático |
| LDA | Lineal Discriminant Analysis | Análisis discriminante lineal |
| FTR | Full Time Result | Resultado al final del encuentro |
| HTR | Half Time Result | Resultado en la media parte |
| FTHG | Full Time Home Team Goals | Goles del equipo local al final del partido |
| FTAG | Full Time Away Team Goals | Goles del equipo visitante al final del partido |
| Bet365H | Bet365 Home Win | Cuota de la victoria local en Bet365 |
| Bet365D | Bet365 Draw | Cuota de la empate en Bet365 |
| Bet365A | Bet365 Away Win | Cuota de la victoria visitante en Bet365 |
| Bet365 | Casa de apuestas Bet365 | |
| BW | Casa de apuestas Bet&Win | |
| IW | Casa de apuestas InterWetten | |
| LB | Casa de apuestas LadBrokes | |
| PS | Casa de apuestas Pinnacle | |
| VC | Casa de apuestas VC Bet | |
| WH | Casa de apuestas WilliamHill | |

Glosario

Surebet Tipo de apuesta que consiste en apostar los posibles resultados de un evento.

Martingala Estrategia donde la cantidad apostada se dobla en cada iteración.

Esperanza matemática Valor medio de un fenómeno aleatorio, cantidad media que se espera en un experimento.

Prueba de hipótesis Regla que especifica si se puede aceptar o rechazar una afirmación acerca de una población

Análisis de componentes principales Descripción de un set de datos en términos de nuevas variables no correlacionadas

Clusters Grupos

Cluster k-Means Método de agrupamiento que tiene como objetivo agrupar n observaciones en k grupos. Cada observación pertenece al grupo cuyo valor medio es más cercano.

Análisis Discriminante Técnica estadística multivariante que describe las diferencias entre grupos y donde se tienen en cuenta las variables del modelo.

Validación Cruzada Técnica utilizada para garantizar la independencia entre datos de entrenamiento y de prueba.

Regresión Logística Nominal Método de regresión logística para problemas de más de dos posibles resultados discretos

Score Coordenadas de una observación proyectada sobre el plano de componentes principales.

Loading Combinación de los pesos de unas variables correlacionadas.

Histograma Representación de distribuciones de frecuencias.

Diagrama de Pareto Representación que asigna órdenes de prioridad

Dotplot Visualización de la distribución de datos sobre un eje, mostrando frecuencias

Diagrama de Gantt Herramienta gráfica que expone el tiempo de dedicación previsto para diferentes actividades.

Tabla de correspondencia Registro y análisis de la asociación entre dos o más variables

Training Set Conjunto de ejemplos utilizados para ajustar los parametros de un modelo

Test Set Conjunto de ejemplos que se utilizan únicamente para evaluar el rendimiento de un modelo

Capítulo 1

Prefacio

1.1. Origen del proyecto

La estadística es una ciencia que desde que pude empezar a conocerla me ha despertado gran interés y hambre de seguir aprendiendo. Anteriormente a mi etapa universitaria ya tuve la ocasión de trabajar en ello. Realicé un proyecto de recerca titulado “La probabilitat als jocs d’atzar”, donde se comenzó por conocer aquellos conceptos y términos necesarios, se analizó la probabilidad de ganar en un largo surtido de juegos de azar, y se llegó a unas conclusiones acerca de la viabilidad de cada uno. Con estos conocimientos, se optó por escoger un trabajo de investigación titulado “Estudi sobre els diferents mètodes per guanyar a la ruleta”, que se focalizó bastante en un estudio físico-probabilístico que acotaba el rango dónde podía caer la bola.

Con esta experiencia, junto con las herramientas de análisis de datos y conocimientos alcanzados durante el grado, la curiosidad por resolver incógnitas y la oportunidad de desarrollarme y aprender, resultan en la realización de éste nuevo proyecto.

1.2. Motivación

Haber trabajado en proyectos de la misma temática durante etapas educativas inferiores motiva a volver a descubrir sobre el tema, esta vez en un nivel superior de exigencia.

Con la idea de realizar un máster de esta rama el curso siguiente, comenzar a desarrollarme con estas materias, y poniéndolo en práctica, es un punto a favor para haber escogido este proyecto.

Resolver la incógnita personal de si se pueden establecer unas estrategias para obtener beneficios, y en mi caso, de si con unos buenos datos y herramientas seré capaz de alcanzar buenos resultados.

Capítulo 2

Introducción

2.1. Objetivos del proyecto

El foco principal de todo el desarrollo del proyecto será conocer al menos una estrategia para apostar en partidos de fútbol de *LaLiga*, y una vez conseguida evaluar su viabilidad económica. Para alcanzar este objetivo habrá que ir alcanzando otros objetivos por el camino, como son el dar respuesta a las cuestiones planteadas en el apartado “preguntas aún sin respuesta”.

Aprender nuevas técnicas estadísticas o de análisis de datos, tanto su base teórica como sus procedimientos y evaluaciones, se encuentra como un objetivo personal encarado al aprendizaje.

Aplicar cualquier conocimiento adquirido durante el grado, como son los aportados explícitamente por las materias de Estadística, Técnicas para la Calidad, Análisis de Datos para la Industria y los Servicios, Álgebra, Geometría, Informática, Organización y Gestión, Gestión de Proyectos, Análisis de Datos para la Industria y los Servicios, así como también de manera implícita por medio de las capacidades desarrolladas gracias al resto, y a cualquier informe realizado en ellas.

Una vez acabado el proyecto, se debería poder compartir la información de la estrategia a una tercera persona, que desconozca tanto de la totalidad del proyecto como las apuestas donde apostará, y ser capaz de obtener una rentabilidad.

2.2. Alcance del proyecto

Como se viene comentando, como mínimo se exige establecer una estrategia. A partir de este punto, la adición de otras estrategias será valorado positivamente.

Las estrategias únicamente serán aplicables a temporadas de la liga española, y desde el proyecto se aconseja únicamente aplicarlos a la temporada posterior de la que se extrajo el modelo.

Solo se trabajará intentando alcanzar beneficios en apuestas del tipo 1-X-2, y nunca utilizando las cuotas de durante el partido. En el proyecto se analizará la primera división de la liga española.

El programario a utilizar será Minitab, Microsoft Excel y TeXstudio.

Capítulo 3

Apuestas y cuotas

3.1. Breve introducción a las apuestas

Desde el siglo XX las apuestas están presentes en la sociedad. No siempre el objetivo es puramente el de obtener unas ganancias, ni el método de operar totalmente racional. Simplemente pueden ser un modo de ocio, entretenimiento o pasatiempo. Se pueden distinguir diversos juegos de azar donde apostar. Para el interés del proyecto se realiza la siguiente clasificación:

- **Juegos de azar puro:** Consisten en una repartición de participaciones donde únicamente una parte de ellos serán premiados. La única variable en juego es el azar. Existe algún intermediario que obtiene un beneficio. Por ejemplo, las loterías un 30 % de lo recaudado recae en dicho intermediario, y en el caso de la ruleta americana $1/36$ recae en el casino.
- **Resto:** A parte de una componente de azar, existen otras variables en juego, controlables o no. Ejemplos serían los juegos de cartas o las apuestas deportivas.

Ambas distinciones comparten la presencia del intermediario mencionado. Por lo tanto, si se observa desde un punto de vista matemático, la participación en juegos de azar puro resulta inútil, por ligera que sea la ventaja del intermediario. En el resto de juegos, entra en juego el factor conocimiento, experiencia o incluso estrategias para apostar.

Por lo tanto, en este punto aún no se puede afirmar la inviabilidad de tratar con apuestas deportivas.

3.1.1. Preguntas aún sin respuesta

En concreto, para el caso de apuestas deportivas de fútbol, se realizaron las siguientes preguntas:

- ¿Las casas de apuesta tienen un margen de seguridad, o parten con una cierta ventaja en sus cuotas? Dado que es previsible que así sea, ¿qué ventaja tienen?

- Partiendo de la premisa de que hay equipos que acostumbran a ganar un considerable porcentaje de sus partidos, ¿es rentable tomar como estrategia apostar durante un periodo siempre a favor de éste?, si no fuese así, ¿una estrategia a largo plazo podría ser apostar siempre en contra del favorito?
- Dado que las cuotas son parcialmente resultado de a qué y cuánto el participante a apostado, ¿es posible que éste no apueste bien, esto sea detectable y se pueda aprovechar? Con otras palabras, ¿son las cuotas reflejo de las probabilidades?
- La existencia de varias casas de apuestas, cada una con sus cuotas, puede parecer interesante, ¿realmente puede pasar a ser una ventaja?
- ¿Existe alguna estrategia para obtener beneficio?
- ¿Se podría llegar a crear un modelo predictivo, donde participen varias variables, y que se ajuste suficientemente al resultado?
- ¿Existe algún tipo de encuentro preferible para apostar?

3.2. Tema a tratar

Hasta el momento se ha comentado que se analizarán apuestas deportivas, y en concreto de fútbol.

Para poder hacer un análisis más exhaustivo se ha tomado la decisión de limitar el estudio a la primera división de la liga española. Los datos corresponden a la temporada 2016/17. Nunca se intentará predecir la cantidad de goles en un partido, tiros a puerta, resultados exactos ni ninguna otra característica del partido que no sea el resultado final del encuentro, entendiéndose como si el vencedor resulto ser el equipo local, el visitante, o el partido acabó en empate.

En esta liga participan 20 equipos, y durante el desarrollo de la temporada cada equipo se enfrentará a los 19 restantes, y lo hará tanto de local como de visitante. Por lo tanto en cada jornada de la temporada habrá 10 partidos, y una temporada tendrá 38 jornadas.

3.3. Fuente de datos

Los análisis se realizarán a partir de unas tablas de datos extraídas de una web [1] que ofrece predicciones a sus subscriptores, y al resto les permite el acceso a algunos de sus datos. Se dispone de ficheros de 10 países europeos, con una o más división por país, y para cada una de ellas una recogida de datos de más de 15 temporadas.

Cada fichero contiene información suficiente para abarcar el proyecto. Para los datos a utilizar se dispone de:

- Filas: 380, cada una de ellas corresponde a un partido. Están ordenadas temporalmente.
- Columnas: 64. En ellas se describe el partido en juego, el resultado del partido, las cuotas del partido para diferentes circunstancias de 7 casas de apuestas diferentes. También se dispone de otros datos de menor interés para el proyecto, como cantidad de remates a puerta, tarjetas, fueras de juego o corners, entre otros.

En el apartado 3.4, “Contenido del fichero”, se describe la información que contiene de manera más detallada. Los datos referidos a las cuotas son recopilados cada viernes por la tarde, para los encuentros de fin de semana, y cada martes por la tarde, para los encuentros de entre semana. El fichero utilizado se muestra en el anexo A, el cual incorpora las figuras A.1, A.2 y A.3.

3.4. Contenido del fichero

- Equipos y resultado del partido:

| | |
|-----------------|--|
| Div | División |
| Date | Fecha del partido (dd/mm/aa) |
| HomeTeam | Nombre del equipo local |
| AwayTeam | Nombre del equipo visitante |
| FTHG | Goles marcados por el equipo local durante el partido |
| FTAG | Goles marcados por el equipo visitante durante el partido |
| FTR | Resultado al final del encuentro |
| HTHG | Goles marcados por el equipo local en la primera mitad |
| HTAG | Goles marcados por el equipo visitante en la primera mitad |
| HTR | Resultado al final de la primera parte |

■ Estadísticas del partido:

| | |
|------------|---|
| HS | Tiros a puerta del equipo local |
| AS | Tiros a puerta del equipo visitante |
| HST | Tiros a puerta del equipo local |
| AST | Tiros a puerta del equipo visitante |
| HHW | Tiros al palo del equipo local |
| AHW | Tiros al palo del equipo visitante |
| HC | Córners a favor del equipo local |
| AF | Córners a favor del equipo visitante |
| HO | Fueras de juego del equipo local |
| AO | Fueras de juego del equipo visitante |
| HY | Tarjetas amarillas para el equipo local |
| AY | Tarjetas amarillas para el equipo visitante |
| HR | Tarjetas rojas para el equipo local |
| AR | Tarjetas rojas para el equipo visitante |

■ Cuotas del partido:

El último carácter será H,D o A, refiriéndose a si la cuota del partido es la pagada en caso de ganar el equipo local, de si hubiera empate, o de victoria visitante. Los caracteres anteriores son siglas distintivas de la casa de apuestas. A continuación sólo se mostrarán los diferentes tipos de apuestas para el caso de la primera casa de apuestas, y el resto es repetición del mismo tipo de apuestas para el resto de casas. Muestra de a qué casa de apuestas corresponden las diferentes siglas:

| | |
|------------|---|
| HS | Tiros a puerta del equipo local |
| AS | Tiros a puerta del equipo visitante |
| HST | Tiros a puerta del equipo local |
| AST | Tiros a puerta del equipo visitante |
| HHW | Tiros al palo del equipo local |
| AHW | Tiros al palo del equipo visitante |
| HC | Córners a favor del equipo local |
| AF | Córners a favor del equipo visitante |
| HO | Fueras de juego del equipo local |
| AO | Fueras de juego del equipo visitante |
| HY | Tarjetas amarillas para el equipo local |
| AY | Tarjetas amarillas para el equipo visitante |
| HR | Tarjetas rojas para el equipo local |
| AR | Tarjetas rojas para el equipo visitante |

3.5. Funcionamiento de las cuotas

Saber qué son las cuotas y cómo fluctúan es una de las bases para entender lo que se va a realizar. Como ya se ha comentado, las apuestas con las que se trabajarán son las de tipo 1X2. En la figura 3.1 se muestran las cuotas correspondientes a las dos primeras jornadas de la temporada a tratar:

| | Home Team | Away Team | B365H | B365D | B365A |
|------------|-------------|-------------|-------|-------|-------|
| 19/08/2016 | La Coruna | Eibar | 1,95 | 3,4 | 4,2 |
| 19/08/2016 | Malaga | Osasuna | 1,62 | 3,75 | 6 |
| 20/08/2016 | Barcelona | Betis | 1,08 | 11 | 29 |
| 20/08/2016 | Granada | Villarreal | 2,5 | 3,2 | 3 |
| 20/08/2016 | Sevilla | Espanol | 1,67 | 4 | 5 |
| 21/08/2016 | Ath Madrid | Alaves | 1,2 | 6,5 | 17 |
| 21/08/2016 | Sociedad | Real Madrid | 4,75 | 3,8 | 1,75 |
| 21/08/2016 | Sp Gijon | Ath Bilbao | 4 | 3,5 | 1,95 |
| 22/08/2016 | Celta | Leganes | 1,65 | 3,75 | 5,75 |
| 22/08/2016 | Valencia | Las Palmas | 1,73 | 3,75 | 5 |
| 26/08/2016 | Betis | La Coruna | 2,2 | 3,2 | 3,5 |
| 26/08/2016 | Espanol | Malaga | 2,3 | 3,1 | 3,4 |
| 27/08/2016 | Eibar | Valencia | 2,7 | 3,2 | 2,75 |
| 27/08/2016 | Leganes | Ath Madrid | 8,5 | 3,6 | 1,53 |
| 27/08/2016 | Osasuna | Sociedad | 3,1 | 3,1 | 2,5 |
| 27/08/2016 | Real Madrid | Celta | 1,18 | 7 | 17 |
| 28/08/2016 | Alaves | Sp Gijon | 2,15 | 3,2 | 3,75 |
| 28/08/2016 | Ath Bilbao | Barcelona | 6,5 | 4,33 | 1,5 |

FIGURA 3.1: Captura de las cuotas tipo 1X2 de las dos primeras jornadas

Para el primer partido que encontramos, la casa de apuestas Bet365, ofrece unas cuotas de 1.95, 3.4 y de 4.2 para la victoria local, empate y victoria visitante respectivamente. Las cuotas indican la ponderación que obtendría una cierta cantidad apostada en el caso de cumplirse el pronóstico. Es decir, que si el apostante confía en la victoria del Eibar y deposita una cantidad X, la casa de apuestas le pagaría $4.2 \cdot X$ en caso de que Eibar ganase.

| Equipo local | Equipo visitante | Cuota local | Cuota empate | Cuota visitante |
|--------------|------------------|-------------|--------------|-----------------|
| La Coruña | Eibar | 1.95 | 3.4 | 4.2 |

CUADRO 3.1: Ejemplo cuotas

Ejemplo de posibles casos:

| Apostar a | Cantidad postada (CA) | Vencedor hipotético | Ganancia (G) | Balance (GN=G·CA) |
|------------------|------------------------------|----------------------------|-----------------------|------------------------------|
| La Coruña | 1 | La Coruña | $1 \cdot 1.95 = 1.95$ | 0.95 |
| La Coruña | 2 | La Coruña | $2 \cdot 1.95 = 3.9$ | 1.9 |
| La Coruña | 1 | Empate | 0 | 0.1 |
| Empate | 1 | La Coruña | 0 | 0.1 |
| Empate | 2 | Empate | $2 \cdot 3.4 = 6.8$ | 4.8 |
| Empate | 1 | Eibar | 0 | .1 |
| Eibar | 1 | La Coruña | 0 | .1 |
| Eibar | 2 | Empate | 0 | .2 |
| Eibar | 1 | Eibar | $1 \cdot 4.2 = 4.2$ | 3.2 |

CUADRO 3.2: Ejemplo de unas aplicaciones de cuotas para casos hipotéticos

Capítulo 4

Primeros análisis

4.1. Resultado final

Hacer una ligera observación sobre el terreno puede ayudar a hacerse una idea de las tendencias, o incluso a hacerse preguntas que pueden ser interesantes de resolver. Se comienza por ver qué resultado se ha producido más. De los 380 partidos de la temporada a estudiar, casi en la mitad de ellos el equipo local resultó vencedor, como se observa en la figura 4.1.

El diagrama de Pareto es una manera de reflejar orden de prioridades. A su izquierda muestra los grupos que más se repitieron, y moviéndose hacia la derecha aquellos de menor repetición.

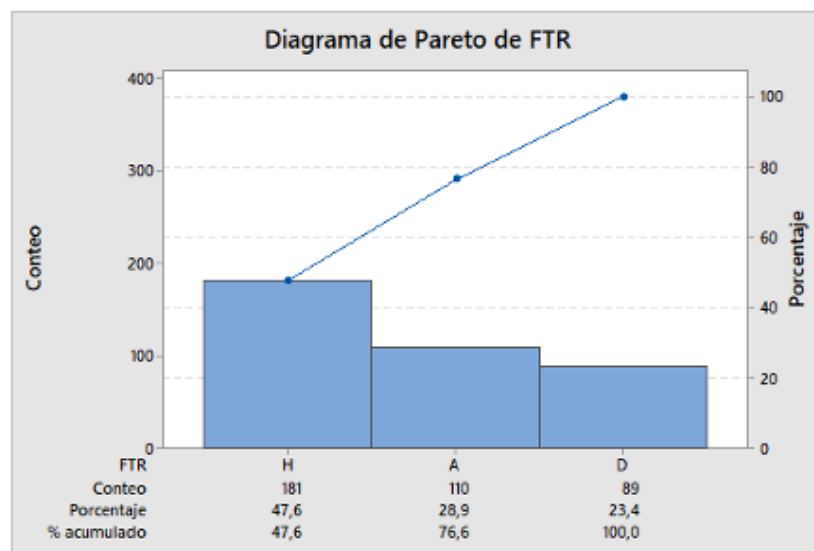


FIGURA 4.1: Diagrama de Pareto de los resultados de final de partido

En cambio, el resultado que más se repetía en el descanso del partido era el de empate, como se observa en la figura 4.2.

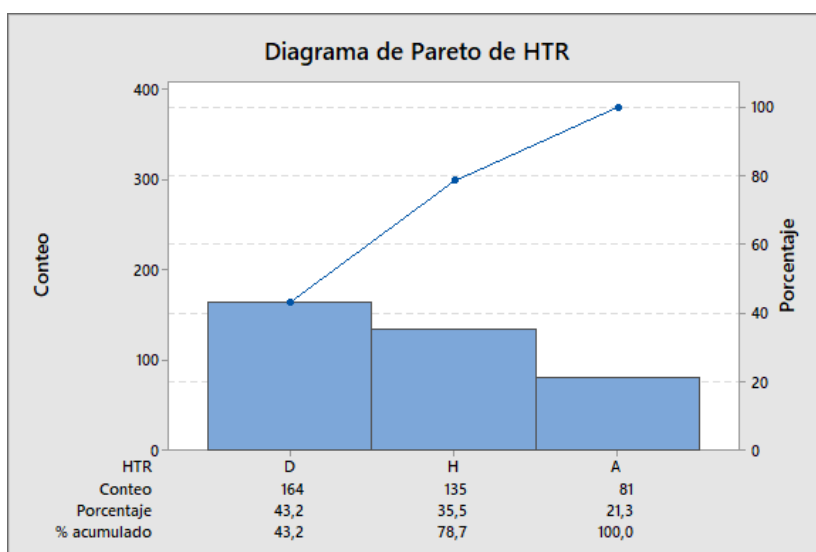


FIGURA 4.2: Diagrama de Pareto de los resultados en la media parte

Entonces, ¿qué relación existe entre resultados en el descanso del partido y al final del partido? En la figura 4.3 se muestra como cuando en el descanso el equipo vencedor es el local, la mayoría de partidos acaban finalizando con victoria local.

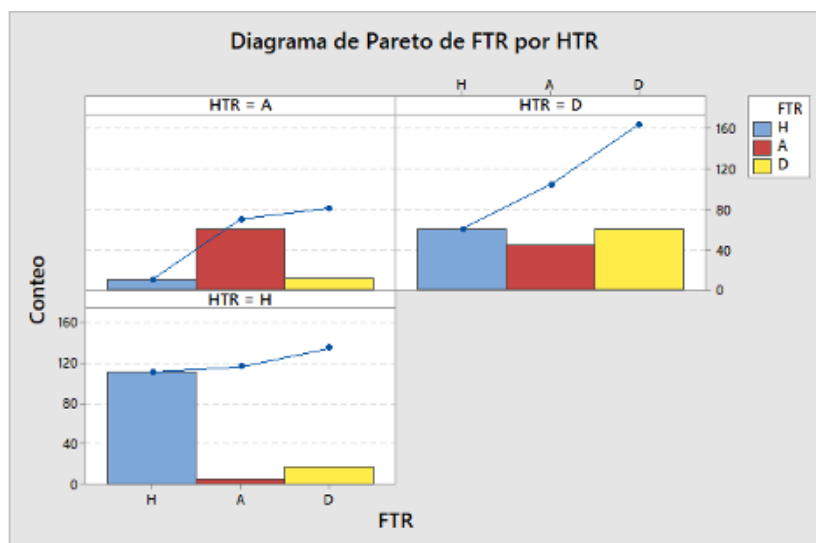


FIGURA 4.3: Diagrama de Pareto de los resultados finales estratificados según resultado al descanso

- Con este pequeño análisis ya surge una primera incógnita, de si resultaría rentable apostar a favor del equipo local cuando éste ya va por delante en el descanso. En este caso no se resolverá, pues se aleja del objetivo del proyecto, donde únicamente se pretenden utilizar las cuotas de antes del partido y no las de durante. Sin embargo, sí que sirve para encaminar a que mediante la exploración de datos se pueden ver patrones, y se espera que en el desarrollo del trabajo algunos de ellos puedan evaluarse.

4.2. Favorito

Muchos son los factores que influyen en que un equipo sea favorito. Algunos son subjetivos y otros son más comunes para los apostantes. Uno de los objetivos finales del proyecto es determinar, en base a datos, qué factores son determinantes para conocer quién ganará, que viene a ser lo que aquí se define como favorito.

Algunos factores de peso comúnmente asociados al favorito pueden ser: jugar como local, tendencias ganadoras del equipo, los días de descanso respecto al último partido, el poder disponer de sus mejores jugadores y que estén en buen estado físico, o el rival contra el que se enfrenta.

Dado que es inviable realizar una búsqueda de esta categoría de datos y además el proyecto va en otra dirección, simplemente se fijó como favorito aquel equipo de cuota inferior. En la tabla 4.1 se presentan las cuotas de un partido de la primera jornada, y que según el criterio definido, el Barcelona fue el favorito. Intuitivamente, y bajo la idea que jugar como local es un factor positivo, se puede pensar que en más ocasiones el favorito es el local, y no el visitante. Se probó y efectivamente así fue, en la figura 4.4 se muestra una captura con varios ejemplos, y en la figura 4.5 el recuento, que resulta en que en el 72 % de los partidos de la temporada de estudio el equipo local partía como favorito.

| Equipo local | Equipo visitante | Cuota local | Cuota empate | Cuota visitante |
|--------------|------------------|-------------|--------------|-----------------|
| Barcelona | Betis | 1.08 | 11 | 29 |

CUADRO 4.1: Ejemplo de la consideración de favorito

| Home Team | Away Team | FTHG | FTAG | FTR | B360H | B360D | B360A | Favorito |
|-------------|-------------|------|------|-----|-------|-------|-------|----------|
| La Coruña | Elbar | 2 | 1 | H | 1.95 | 3.4 | 4.2 | H |
| Malaga | Osasuna | 1 | 1 | D | 1.62 | 3.75 | 6 | H |
| Barcelona | Betis | 6 | 2 | H | 1.08 | 11 | 29 | H |
| Granada | Villarreal | 1 | 1 | D | 2.5 | 3.2 | 3 | H |
| Sevilla | Espanol | 6 | 4 | H | 1.57 | 4 | 5 | H |
| Ath Madrid | Alaves | 1 | 1 | D | 1.2 | 6.5 | 17 | H |
| Sociedad | Real Madrid | 0 | 3 | A | 4.75 | 3.8 | 1.75 | A |
| Sp Gijon | Ath Bilbao | 2 | 1 | H | 4 | 3.5 | 1.95 | A |
| Celta | Leganes | 0 | 1 | A | 1.55 | 3.75 | 5.75 | H |
| Valencia | Las Palmas | 2 | 4 | A | 1.73 | 3.75 | 5 | H |
| Betis | La Coruña | 0 | 0 | D | 2.2 | 3.2 | 3.5 | H |
| Espanol | Malaga | 2 | 2 | D | 2.3 | 3.1 | 3.4 | H |
| Eibar | Valencia | 1 | 0 | H | 2.7 | 3.2 | 2.75 | H |
| Leganes | Ath Madrid | 0 | 0 | D | 8.5 | 3.6 | 1.53 | A |
| Osasuna | Sociedad | 0 | 2 | A | 3.1 | 3.1 | 2.5 | A |
| Real Madrid | Celta | 2 | 1 | H | 1.18 | 7 | 17 | H |
| Alaves | Sp Gijon | 0 | 0 | D | 2.15 | 3.2 | 3.75 | H |
| Ath Bilbao | Barcelona | 0 | 1 | A | 6.5 | 4.33 | 1.5 | A |

FIGURA 4.4: Captura de las cuotas tipo 1X2 y de quién es favorito.

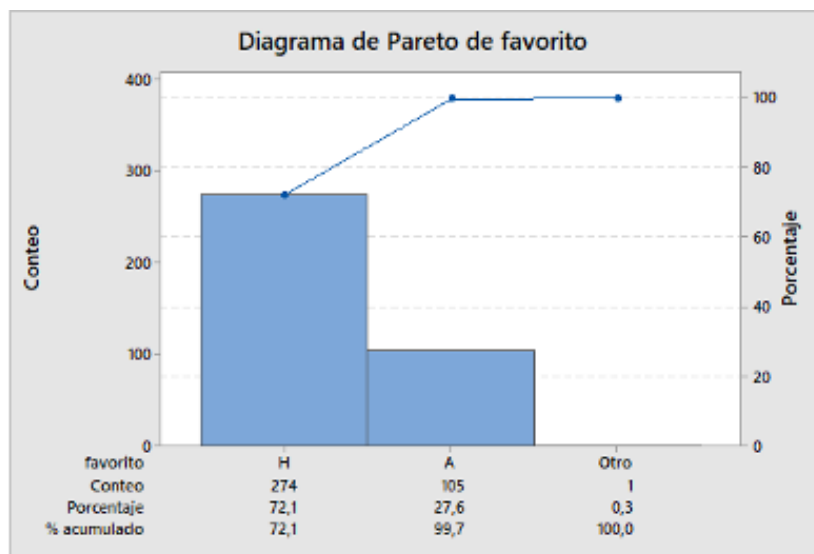


FIGURA 4.5: Diagrama de Pareto de qué equipo fue favorito

Dado que el porcentaje de partidos en que el favorito es el local es considerable, se procedió a ver cuál fue el resultado final únicamente en estos partidos de local favorito. En la figura 4.6 se observa como en un 58.8 % de los partidos donde el local era favorito, acabó resultando ganador. Para el caso en que el favorito es el visitante, también se procedió a mostrar sus resultados finales, tal y como se pueden ver en el gráfico 4.7, y el porcentaje de acierto es bastante similar, de un 59 %.

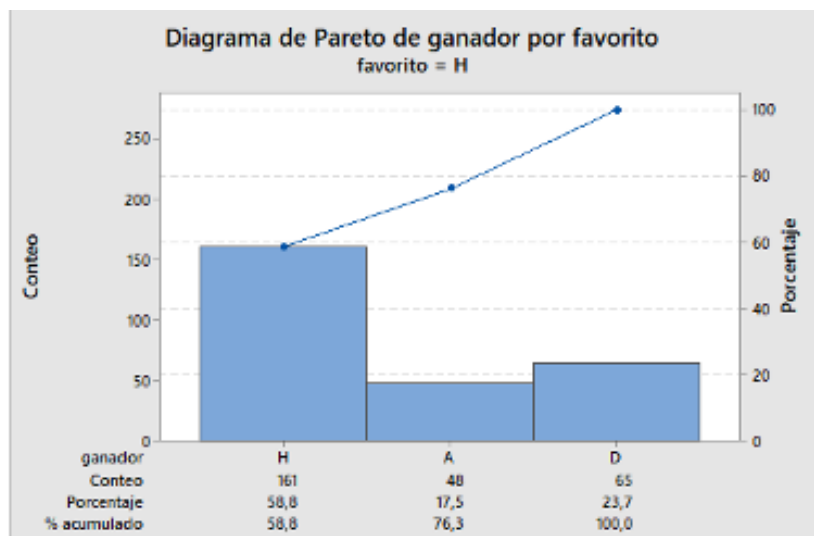


FIGURA 4.6: Diagrama de Pareto de resultado final cuando el favorito es el local

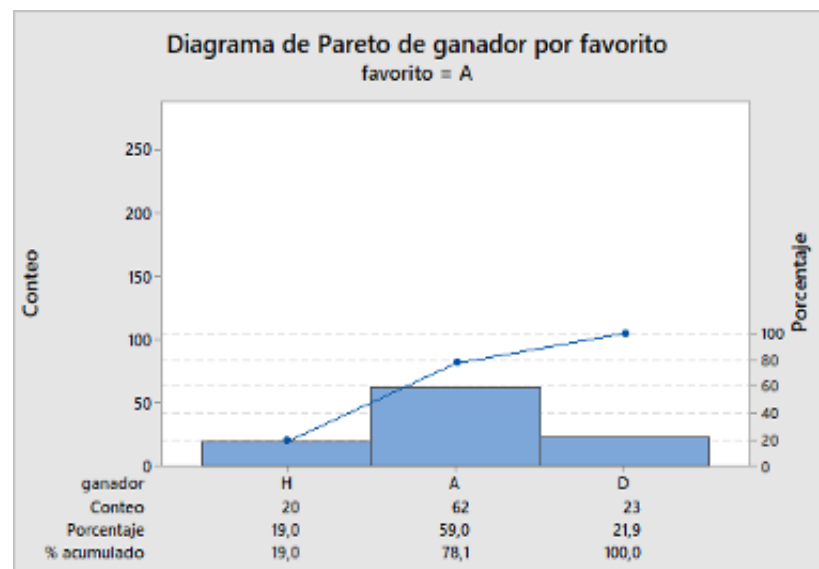


FIGURA 4.7: Diagrama de Pareto de resultado final cuando el favorito es el visitante

En ningún caso la condición de favorito asegura más del 60 % de aciertos. Si las cuotas de favorito cubren las pérdidas resultaría viable tomar por estrategia apostar al equipo favorito.

Se realiza una tabla donde se ve representado cuándo el ganador era el favorito, cuál era la cuota que se pagó por el resultado correcto y cuál fue la cuota obtenida apostando al favorito.

Seguidamente, se procede a realizar una prueba con todos los partidos de la temporada, apostando siempre al equipo favorito. El resultado es una apuesta que ofrece beneficios, tal y como se muestra en la tabla 4.2, todo y que como se muestra en la figura 4.8, según una prueba de hipótesis, donde se valora si el valor medio de ganancia podría ser inferior a la unidad, resulta que la estrategia no se puede confirmar como segura para probarla en otras temporadas. La prueba de hipótesis es una metodología de inferencia estadística que permite juzgar si en una población una propiedad hipotéticamente cierta es también así bajo la observación de una muestra de dicha población.

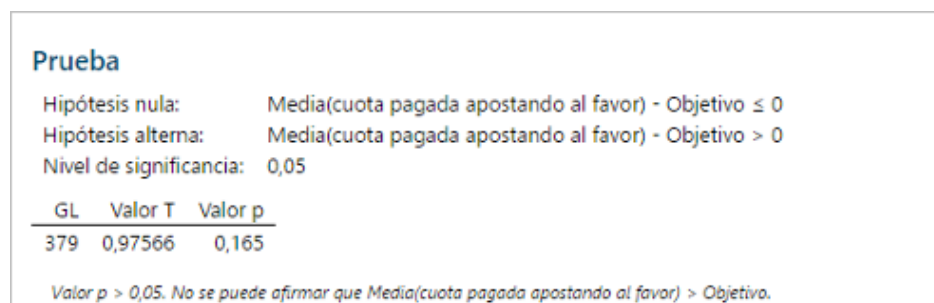


FIGURA 4.8: Captura de Minitab de la prueba de hipótesis para apuestas al favorito

| Cantidad apostada por partido | Cantidad apostada en la temporada | Ganancias | Balance | Estrategia segura |
|----------------------------------|--------------------------------------|-----------|---------|----------------------|
| 1€ | 380€ | 398,13€ | +18,13€ | NO |

CUADRO 4.2: Resultado de apostar toda una temporada al favorito

Campeón de la temporada: El Real Madrid fue el campeón de la temporada con la que se trabaja. ¿Qué hubiera pasado en caso de haber apostado una misma cantidad en todos sus encuentros de la temporada, por ejemplo, en condición de local? En la figura 4.9 se muestran los valores de las cuotas únicamente en aquellos partidos en los que ganó, y en la tabla 4.3 se muestra qué hubiera pasado en caso de apostar a favor dicho equipo y en condición de local.

Se ha escogido al equipo campeón de liga porque al haber resultado campeón, se podía pensar que apostar por él implica ganancia. En la línea del objetivo del proyecto, asegurar un elevado número de aciertos no implica obtener beneficios económicos. En este ejemplo, resulta que las cuotas del Real Madrid como local son tan bajas que no son capaces de cubrir la posibilidad de no ganar todos sus partidos.

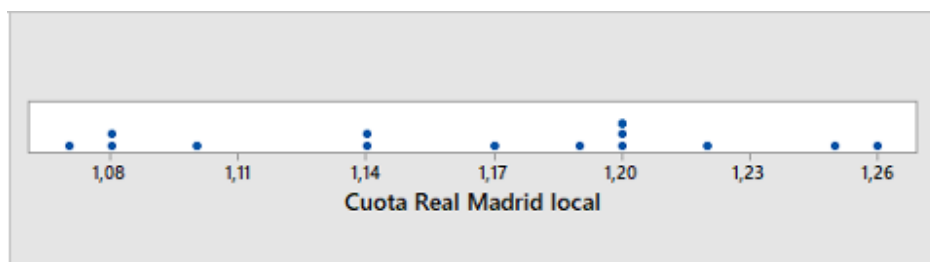


FIGURA 4.9: Dotplot de las cuotas del Real Madrid cuando fue local y además ganó

| Cantidad apostada por partido | Cantidad apostada como local | Ganancias | Balance | Estrategia segura |
|----------------------------------|---------------------------------|-----------|---------|----------------------|
| 1€ | 19€ | 16.3€ | -2.7€ | NO |

CUADRO 4.3: Resultado de apostar a un favorito toda la temporada

- Llegados a este punto, se comienza a ver que las casas de apuestas manejan correctamente el valor de sus cuotas y ante métodos de apostar tan simples ya están cubiertas. En adelante, los métodos no entenderán de quién participa en el partido y se alejarán del mayor o menor conocimiento del apostante sobre este deporte y sobre esta liga.

4.3. Transformación del valor de cuota a probabilidad de éxito

Si de alguna manera se pudiera asegurar el correcto valor de las cuotas, asociado a la probabilidad de éxito de la apuesta a la que pertenece, también se podría obtener la probabilidad de éxito que la apuesta tiene. En la tabla 4.4 se muestran las cuotas de un partido de la primera jornada, en la ecuación 4.1 se muestra el cambio a realizar y en la tabla 4.5 el cuadro de probabilidades.

| Equipo local | Equipo visitante | Cuota local | Cuota empate | Cuota visitante |
|--------------|------------------|-------------|--------------|-----------------|
| Valencia | Las Palmas | 1.73 | 3.75 | 5 |

CUADRO 4.4: Ejemplo para transformar el valor de cuota a probabilidad de éxito

$$\text{Probabilidad éxito } (X) = \frac{1}{\text{Cuota } (X)} \quad (4.1)$$

| Resultado | Cuota | Probabilidad éxito |
|---------------------|-------|--------------------|
| Gana Valencia | 1.73 | 0.578 |
| Empate | 3.75 | 0.267 |
| Gana Las Palmas | 5 | 0.2 |
| Cualquier situación | - | 1.045 |

CUADRO 4.5: Cuotas y probabilidades.

Si se suma la probabilidad de éxito de cada situación posible no suma el valor intuitivamente esperado, la unidad. Que el valor sea superior a la unidad viene a decir que probabilísticamente la ventaja la tiene la casa de apuestas.

Inventando valores para una nueva situación ficticia se podrá entender que al jugador no le interesa que la suma final sea superior a la unidad. En la tabla 4.6 se muestran cuotas ficticias superiores a las reales, y se ve como suma de probabilidades de éxito es menor a la unidad. También se muestran cuotas ficticias inferiores y la probabilidades de éxito es superior a la unidad.

Inciso:

- En la ruleta americana de casino la casa de apuestas vuelven a tener una ventaja. Si se apuesta a un número aleatorio, como por ejemplo el 32, y se acierta, la casa pagará 36 veces la apuesta. La cuota es de 36 cuando debería ser de 37. Pues hay 37 casillas, ya que existe una con el valor de 0.
- En la ruleta americana de doble cero la banca aún tiene más ventaja, pues la probabilidad de éxito pasa a ser 1/38 en vez de 1/37, sin embargo la cuota se mantiene en 36.
- Tanto las ruletas como el ejemplo visto del encuentro, no son juegos de suma cero. Las ganancias y pérdidas entre participantes no se equilibran, existe el ya mencionado intermediario beneficiado.

Factor corrector: Si se quisiera conocer la probabilidad real de éxito, conocida a partir de las cuotas, debe incorporarse un factor corrector que fuerce la suma de probabilidades a la unidad. En la ecuación 4.2 se muestra como se calcularía las nuevas probabilidades de éxito, y en la tabla 4.6 el ejemplo mostrado en la tabla 4.4, pero con las probabilidades de éxito corregidas.

$$Prob. \text{ éxito corregida}(X) = \frac{\frac{1}{Cuota(X)}}{\sum_{i=1}^n \frac{1}{Cuota(i)}} \text{ para } n = 1, X, 2 \quad (4.2)$$

| Resultado | Cuota | Probabilidad éxito | Probabilidad éxito corregida |
|---------------------|-------|--------------------|------------------------------|
| Gana Valencia | 1.73 | 0.578 | 0.553 |
| Empate | 3.75 | 0.267 | 0.256 |
| Gana Las Palmas | 5 | 0.2 | 0.191 |
| Cualquier situación | - | 1.045 | 1.000 |

CUADRO 4.6: Probabilidades de éxito corregidas

- Es importante conocer la probabilidad corregida y será necesario para una estrategia aún no presentada, además de por lo que implica que las probabilidades sumen mayor a la unidad.

Capítulo 5

Estrategias

5.1. Estrategias simples

Previo a la búsqueda del diseño una estrategia se realizan unas pruebas con estrategias muy básicas para ver cuál es la situación inicial, así posteriormente se podrá valorar la mejora en el momento de utilizar estrategias más sofisticadas. Las estrategias más básicas se muestran en las tablas 5.1, 5.2 y 5.3, y corresponden a haber evaluado las apuestas de siempre apuesta al local, siempre apuesta al visitante y siempre apuesta al empate, respectivamente. En cada tabla se muestra que las estrategias se han probado con cada casa de apuestas diferente. Las esperanzas matemáticas varían porque los valores de las cuotas en cada casa de apuestas es diferente. El porcentaje de aciertos para cada estrategia debería ser siempre el mismo, pero si en algún partido la base de datos no recogía las cuotas de una casa concreta, pues obviamente no se pudo apostar.

En las tablas 5.4 y 5.5 ya se muestran unas estrategias básicas pero con un añadido, y es que únicamente se apostó al local cuando éste era favorito, y de igual manera con el visitante.

En la tabla 5.6 se recogen los resultados de las 5 estrategias básicas mencionadas. En la continuación del proyecto la idea es alcanzar una o más de una estrategia que supere estos resultados, sobre todo el de esperanza matemática.

| Casa de apuestas | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Media |
|-------------------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Esperanza matemática | 0,961 | 0,953 | 0.956 | 0.940 | 0.984 | 0.966 | 0.967 | 0.961 |
| Porcentaje aciertos | 47.63 | 47.63 | 47.10 | 47.36 | 47.63 | 47.63 | 47.63 | 47.51 |

CUADRO 5.1: Apuesta simple al local (L).

| Casa de apuestas | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Media |
|-----------------------------|----------|----------|----------|----------|----------|----------|----------|--------------|
| Esperanza matemática | 0.859 | 0.859 | 0.807 | 0.855 | 0.893 | 0.840 | 0.878 | 0.856 |
| Porcentaje aciertos | 28.94 | 28.04 | 28.68 | 28.94 | 28.94 | 28.94 | 28,94 | 23.77 |

CUADRO 5.2: Apuesta simple al visitante (V).

| Casa de apuestas | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Media |
|-----------------------------|----------|----------|----------|----------|----------|----------|----------|--------------|
| Esperanza matemática | 0.917 | 0.915 | 0.910 | 0.898 | 0.965 | 0.899 | 0.952 | 0.922 |
| Porcentaje aciertos | 23.42 | 23.42 | 23.42 | 23.42 | 23.42 | 23.42 | 23.42 | 23.42 |

CUADRO 5.3: Apuesta simple al empate (E).

| Casa de apuestas | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Media |
|-----------------------------|----------|----------|----------|----------|----------|----------|----------|--------------|
| Esperanza matemática | 0.985 | 0.987 | 0.957 | 0.991 | 1.016 | 1.008 | 0.999 | 0.992 |
| Porcentaje aciertos | 59.04 | 58.69 | 58.23 | 59.25 | 59.19 | 59.19 | 58.97 | 58.93 |

CUADRO 5.4: Apuesta simple al local cuando es favorito (L1).

| Casa de apuestas | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Media |
|-----------------------------|----------|----------|----------|----------|----------|----------|----------|--------------|
| Esperanza matemática | 1.031 | 1.039 | 0.926 | 1.055 | 1.097 | 1.050 | 1.035 | 1.033 |
| Porcentaje aciertos | 57.94 | 59.40 | 54.45 | 59.04 | 59.43 | 58.65 | 58.25 | 58.16 |

CUADRO 5.5: Apuesta simple al visitante cuando es favorito (V1).

| Estrategia | L | V | E | L1 | V1 |
|----------------------|-------|-------|-------|-------|-------|
| Esperanza matemática | 0.961 | 0.856 | 0.922 | 0.992 | 1.033 |
| Porcentaje aciertos | 47.71 | 28.77 | 23.42 | 58.93 | 58.16 |

CUADRO 5.6: Resumen de los resultados de cada estrategia simple

Se observa como con las estrategias más simples el valor esperado es inferior a la unidad, lo que dice que no son estrategias factibles. En el momento que se intenta añadir una condición, en este caso ha sido la de comprobar si son favoritos, tanto el porcentaje de aciertos como el valor esperado aumentan.

En las siguientes pruebas se intentará lo mismo, mediante estrategias más sofisticadas hacer crecer estos dos valores. Posteriormente se evaluarán en una temporada diferente, y es por ello que no se debe cometer el error de considerar como estrategias unas en las que en vez de centrarse únicamente en detectar repeticiones lo que haga es sobreajustar. Por ello, la evaluación en una o varias temporadas diferentes será imprescindible.

5.2. Estrategias existentes

A lo largo del capítulo se describirán tres estrategias, y que cada una de ellas podrá tener sus variantes. Para cada estrategia se explicará en qué consiste y se buscará si hubiesen sido viables con los datos disponibles.

5.2.1. Surebet

La traducción literal de Surebet es *apuesta segura*. Éste es un tipo de apuestas que procede de manera sistemática y que actúa únicamente cuando existe posibilidad de emplearla bajo los requisitos que ha de cumplir.

La idea es que dado un tipo de apuesta, como sería el 1X2 en el caso de este proyecto, el apostador deposite diferentes cantidades de dinero. Apostará tanto al equipo local, visitante y empate. La intención está en controlar cuánto apostar a cada situación, tal que dé igual el resultado final, la ganancia sea la misma.

| Resultado | Málaga | Empate | Osasuna |
|-------------------|----------------|----------------|-------------|
| Cuotas | 1.62 | 3.75 | 6 |
| Cantidad apostada | A | B | C |
| Ganancia | $1.62 \cdot A$ | $3.75 \cdot B$ | $6 \cdot C$ |

CUADRO 5.7: Ejemplo de como realizar una Surebet

Se está buscando que independientemente del resultado final, la ganancia se la misma, por lo tanto se debe fijar las cantidades apostadas como se muestra en la ecuación 5.1.

$$Ganancia = 1,62 \cdot A = 3,75 \cdot B = 6 \cdot C \quad (5.1)$$

Es un sistema con un grado de libertad, en el momento en que se escoge el valor de una cuota el resto quedan fijadas. Se desea apostar 1€ a Osasuna, $C=1€$. Por lo tanto, cumpliendo las ecuaciones, $A=3.7$ y $B=1.6$.

En la tabla 5.8 y 5.9 se muestra la aplicación y los resultados de esta técnica para este partido.

| Resultado | Málaga | Empate | Osasuna |
|-------------------|------------------------|------------------------|-------------------|
| Cuotas | 1.62 | 3.75 | 6 |
| Cantidad apostada | 3.7€ | 1.6€ | 1€ |
| Ganancia | $1.62 \cdot 3.7€ = 6€$ | $3.75 \cdot 1.6€ = 6€$ | $6 \cdot 1€ = 6€$ |

CUADRO 5.8: Aplicación de la condición de beneficios iguales en Surebet

| | |
|------------------|---|
| Inversión | $3.7€ + 1.6€ + 1€ = 6.3€$ |
| Ganancia | 6€ |
| Balance | -0.3€ |

CUADRO 5.9: Resultado del ejemplo de Surebet propuesto

En este caso, realizando SureBet no se obtiene beneficio, y en concreto se puede definir el tanto por uno de la rentabilidad como figura en la ecuación 5.2.

$$Rentabilidad = \frac{Ganancia}{Inversión} \quad (5.2)$$

En este caso la rentabilidad es del 0.952. Únicamente interesa aplicar este método en encuentros con rentabilidad superior a la unidad, es decir, que la ganancia sea mayor que la inversión. Para encontrarlos con mayor facilidad, se mostrará la manera rápida de conocer

la rentabilidad a partir de las cuotas. El desarrollo de la demostración se muestra desde las ecuaciones 5.3 y 5.4, hasta la ecuación 5.5, que resume la rentabilidad.

$$A \cdot Cuota\ 1 = B \cdot Cuota\ X = C \cdot Cuota\ 2 = Ganancias \quad (5.3)$$

$$A + B + C = Inversión \quad (5.4)$$

Ajustando las ganancias a la inversión:

$$A \cdot Cuota\ 1 = A + B + C \quad (5.5)$$

$$Cuota\ 1 = 1 + \frac{B}{A} + \frac{C}{A} \quad (5.6)$$

$$\rightarrow \frac{B}{A} = \frac{Cuota\ 1}{Cuota\ X} \quad (5.7)$$

$$\rightarrow \frac{C}{A} = \frac{Cuota\ 1}{Cuota\ 2} \quad (5.8)$$

$$Cuota\ 1 = 1 + \frac{Cuota\ 1}{Cuota\ X} + \frac{Cuota\ 1}{Cuota\ 2} \quad (5.9)$$

Para A=1, la situación es rentable si la ganancia supera a la inversión:

$$Cuota\ 1 > 1 + \frac{Cuota\ 1}{Cuota\ X} + \frac{Cuota\ 1}{Cuota\ 2} \quad (5.10)$$

$$1 > \frac{1}{Cuota\ 1} + \frac{1}{Cuota\ X} + \frac{1}{Cuota\ 2} \quad (5.11)$$

$$Rentabilidad = \frac{Ganancia}{Inversión} = \frac{1}{\frac{1}{Cuota\ 1} + \frac{1}{Cuota\ X} + \frac{1}{Cuota\ 2}} > 1 \quad (5.12)$$

- La rentabilidad corresponde a la inversa del sumatorio de las probabilidades de éxito de las cuotas.

Se procede a calcular la rentabilidad de las cuotas de la casa de apuestas Bet365, y como se muestra en la figura 5.1, ninguna sería rentable con esta estrategia. En la figura 5.2 se representa mediante histograma la frecuencia de rentabilidades, la cual en ningún caso supera la unidad.

| HomeTeam | AwayTeam | B365H | B365D | B365A | Rentabilidad B365 |
|-------------|-------------|-------|-------|-------|-------------------|
| La Coruna | Eibar | 1,95 | 3,4 | 4,2 | 0,9569103279 |
| Malaga | Osasuna | 1,62 | 3,75 | 6 | 0,951818926 |
| Barcelona | Betis | 1,08 | 11 | 29 | 0,9511851767 |
| Granada | Villarreal | 2,5 | 3,2 | 3 | 0,9561783464 |
| Sevilla | Espanol | 1,67 | 4 | 5 | 0,9534706331 |
| Ath Madrid | Alaves | 1,2 | 6,5 | 17 | 0,9560229446 |
| Sociedad | Real Madrid | 4,75 | 3,8 | 1,75 | 0,9568370794 |
| Sp Gijon | Ath Bilbao | 4 | 3,5 | 1,95 | 0,953716155 |
| Celta | Leganes | 1,65 | 3,75 | 5,75 | 0,9554383551 |
| Valencia | Las Palmas | 1,73 | 3,75 | 5 | 0,9572125969 |
| Betis | La Coruna | 2,2 | 3,2 | 3,5 | 0,9498841141 |
| Espanol | Malaga | 2,3 | 3,1 | 3,4 | 0,9510404382 |
| Eibar | Valencia | 2,7 | 3,2 | 2,75 | 0,955557042 |
| Leganes | Ath Madrid | 8,5 | 3,6 | 1,53 | 0,9532706717 |
| Osasuna | Sociedad | 3,1 | 3,1 | 2,5 | 0,9567913047 |
| Real Madrid | Celta | 1,18 | 7 | 17 | 0,9531616372 |
| Alaves | Sp Gijon | 2,15 | 3,2 | 3,75 | 0,9575975792 |
| Ath Bilbao | Barcelona | 6,5 | 4,33 | 1,5 | 0,9510585281 |
| Las Palmas | Granada | 1,91 | 3,4 | 4,33 | 0,9536343003 |

FIGURA 5.1: Captura de tabla con las rentabilidades ofrecidas por Bet365

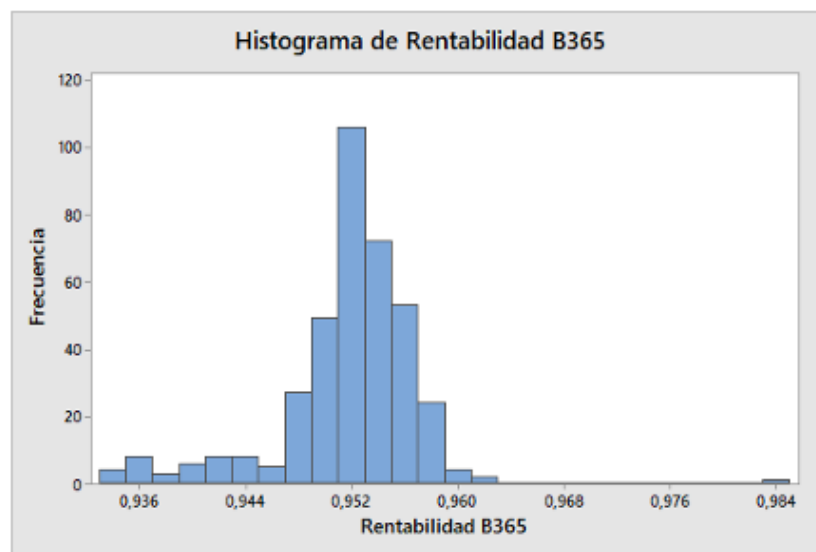


FIGURA 5.2: Histograma de las rentabilidades en Bet365

Se prueba la misma situación para todas las casas de apuestas, y como se muestra en la figura 5.3, tampoco parecen superar la unidad.

| Rentabilidad B365 | Rentabilidad BW | Rentabilidad IW | Rentabilidad LB | Rentabilidad PS | Rentabilidad WH | Rentabilidad VC |
|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 0.9569103279 | 0.9480116135 | 0.9252320019 | 0.9446349458 | 0.9801134971 | 0.9446349458 | 0.9740133245 |
| 0.951818926 | 0.950778212 | 0.9269730622 | 0.9464140372 | 0.9793935595 | 0.9437612662 | 0.9668468225 |
| 0.9511851767 | 0.9507330152 | 0.9224326394 | 0.9589382636 | 0.9804786697 | 0.9323574659 | 0.9580746532 |
| 0.9561783464 | 0.9541347429 | 0.9244451018 | 0.9275061215 | 0.9801519235 | 0.9364172675 | 0.9634004181 |
| 0.9534706331 | 0.9563063623 | 0.925531718 | 0.9324096262 | 0.980545968 | 0.9400705053 | 0.9727342587 |
| 0.9560229446 | 0.9487215976 | 0.9218119135 | 0.9431204082 | 0.9784448597 | 0.9374970703 | 0.9687106461 |
| 0.9568370794 | 0.9541438467 | 0.9253004913 | 0.9462438849 | 0.9795278676 | 0.9403003319 | 0.9650926006 |
| 0.953716155 | 0.95477243 | 0.9252320019 | 0.9366102203 | 0.9804306051 | 0.9427025396 | 0.969847443 |
| 0.9554383551 | 0.9493520672 | 0.925531718 | 0.9448848658 | 0.9822603776 | 0.9417703399 | 0.9734444358 |
| 0.9572125969 | 0.9511218482 | 0.9264663646 | 0.9377871973 | 0.9810846872 | 0.9469158949 | 0.9665664659 |
| 0.9498841141 | 0.9508505358 | 0.9236680706 | 0.9404418196 | 0.9802768302 | 0.9434229271 | 0.9678761893 |
| 0.9510404382 | 0.9557945042 | 0.9253946808 | 0.9389142396 | 0.9794703025 | 0.9478942529 | 0.9647015696 |
| 0.955557042 | 0.94853262 | 0.9244451018 | 0.9305528414 | 0.9808537351 | 0.9421518749 | 0.9723181036 |
| 0.9532706717 | 0.9513118591 | 0.925531718 | 0.9382623381 | 0.9795182729 | 0.9469786646 | 0.9755145839 |
| 0.9567913047 | 0.9561783464 | 0.9244451018 | 0.9470504115 | 0.9812868596 | 0.9374970703 | 0.9698098203 |
| 0.9531616372 | 0.9495684212 | 0.9236424765 | 0.9438236182 | 0.9830618444 | 0.9495684212 | 0.9677450572 |
| 0.9575975792 | 0.9518461055 | 0.9263118892 | 0.9416816551 | 0.9797965942 | 0.9404418196 | 0.9518823473 |
| 0.9510585281 | 0.9510585281 | 0.927179335 | 0.944777741 | 0.9811424423 | 0.9396023603 | 0.9517283387 |
| 0.9536343003 | 0.9522630531 | 0.9266380644 | 0.9384296319 | 0.9806998274 | 0.9461275001 | 0.9518098665 |
| 0.951239465 | 0.952462592 | 0.9255916845 | 0.9432449513 | 0.9797581957 | 0.9432449513 | 0.9528346832 |

FIGURA 5.3: Captura de tablas con rentabilidades en las 7 casas de apuestas

El truco reside en apostar simultáneamente con diferentes casas de apuestas. Para un partido en concreto, localizar aquella casa que mejor pague la victoria de local, la casa que mejor pague el empate y la casa que mejor pague la victoria visitante. Haciendo esto sí que se pondrán ver apuestas de rentabilidad superior a la unidad, como se puede observar en la figura 5.4. En la figura 5.5 se muestra un dotplot con el tanto por uno de los beneficios que se obtendrían en los encuentros con Surebet, y en la figura 5.6 las estadísticas descriptivas de la misma información.

| Mejor cuota local | Mejor cuota empate | Mejor cuota visitante | Existe surebet | Rentabilidad |
|-------------------|--------------------|-----------------------|----------------|--------------|
| 2,01 | 3,4 | 4,39 | 0 | 0,9809499519 |
| 1,7 | 3,89 | 7 | 1 | 1,011981865 |
| 1,1 | 13 | 31 | 0 | 0,9820578039 |
| 2,62 | 3,2 | 3,09 | 0 | 0,9825112969 |
| 1,73 | 4,02 | 5,4 | 0 | 0,9881618214 |
| 1,23 | 6,5 | 20,4 | 0 | 0,9843779224 |
| 5,13 | 3,98 | 1,8 | 0 | 0,9982630223 |
| 4,4 | 3,54 | 2 | 0 | 0,9903343369 |
| 1,65 | 4 | 6,47 | 0 | 0,9894915992 |
| 1,77 | 4 | 5,07 | 0 | 0,9879372857 |
| 2,3 | 3,34 | 3,7 | 0 | 0,9955697148 |
| 2,38 | 3,27 | 3,4 | 0 | 0,9802960494 |
| 2,78 | 3,25 | 2,84 | 0 | 0,9808537351 |
| 8,5 | 3,83 | 1,6 | 0 | 0,9962739355 |
| 3,19 | 3,2 | 2,6 | 0 | 0,9895111815 |
| 1,22 | 7,9 | 17 | 0 | 0,994945676 |
| 2,2 | 3,26 | 4,02 | 0 | 0,9900499975 |
| 6,5 | 4,66 | 1,54 | 0 | 0,9825209523 |
| 2 | 3,6 | 4,58 | 1 | 1,003895113 |

FIGURA 5.4: Captura de tabla con la mejor cuota de cada tipo, y con la nueva rentabilidad

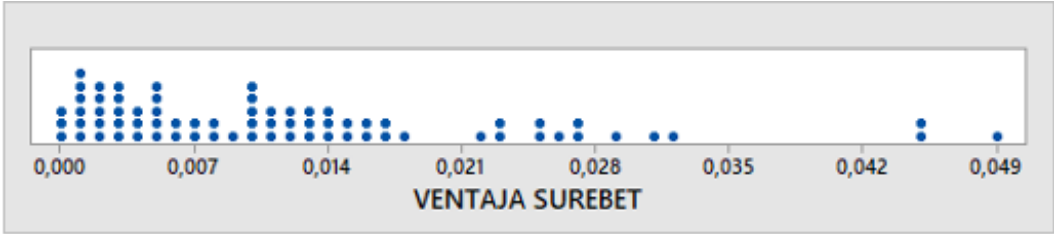


FIGURA 5.5: Dotplot del beneficio neto (en tanto por 1) en presencia de surebet favorable

| Estadísticas | | | | | | | | | |
|-----------------|---------|-----|---------|----------------------------|-----------|---------|---------|---------|---------|
| Variable | N | N* | Media | Error estándar de la media | Desv.Est. | Mínimo | Q1 | Mediana | Q3 |
| VENTAJA SUREBET | 72 | 308 | 0,01200 | 0,00131 | 0,01116 | 0,00013 | 0,00337 | 0,01010 | 0,01631 |
| Variable | Máximo | | | | | | | | |
| VENTAJA SUREBET | 0,04911 | | | | | | | | |

FIGURA 5.6: Captura de las estadísticas descriptivas de la ventaja con surebet

La situación de la temporada 2016-17 fue que de los 380 partidos, en 69 de ellos se podría haber aplicado el sistema SureBet. Realmente no son 72, sino 69, ya que en 3 encuentros la

rentabilidad fue unitaria. En casi uno de cada 5 partidos se pudo participar sin ningún riesgo y con una rentabilidad media superior al 1 %. Para aquel jugador que quiera asegurarse las ganancias es el único método encontrado hasta el momento.

Por intentar descubrir algo más sobre este método aplicado a estos datos, se visualiza qué casas de apuestas suelen tener las mejores cuotas, como también quién acabó efectuando más pagos si se utilizase la apuesta segura. En la figura 5.7 se muestra una captura de cómo se trabajó para obtener ambas informaciones. En las figuras 5.8 hasta la 5.11 se muestran los diagramas de Pareto de la información solicitada.

| Mejor cuota local | Mejor cuota empate | Mejor cuota visitante | MAX H | MAX D | MAX A | Quién ganó | Quién pagó |
|-------------------|--------------------|-----------------------|-------|-------|-------|------------|------------|
| 2.01 | 3.4 | 4.39 | PS | B365 | PS | H | PS |
| 1.7 | 3.89 | 7 | IW | PS | PS | D | PS |
| 1.1 | 13 | 31 | IW | PS | VC | H | IW |
| 2.62 | 3.2 | 3.09 | WH | B365 | PS | D | B365 |
| 1.73 | 4.02 | 5.4 | PS | PS | IW | H | PS |
| 1.23 | 6.5 | 20.4 | PS | B365 | PS | D | B365 |
| 5.13 | 3.98 | 1.8 | PS | PS | WH | A | WH |
| 4.4 | 3.54 | 2 | PS | PS | IW | H | PS |
| 1.65 | 4 | 6.47 | B365 | PS | PS | A | PS |
| 1.77 | 4 | 5.07 | PS | WH | PS | A | PS |
| 2.3 | 3.34 | 3.7 | IW | PS | PS | D | PS |
| 2.38 | 3.27 | 3.4 | PS | PS | B365 | D | PS |
| 2.78 | 3.25 | 2.84 | PS | PS | PS | H | PS |
| 8.5 | 3.83 | 1.6 | B365 | PS | IW | D | PS |
| 3.19 | 3.2 | 2.6 | PS | BW | IW | A | IW |
| 1.22 | 7.9 | 17 | PS | PS | B365 | H | PS |
| 2.2 | 3.26 | 4.02 | IW | PS | PS | D | PS |
| 6.5 | 4.66 | 1.54 | B365 | PS | PS | A | PS |
| 2 | 3.6 | 4.58 | IW | PS | PS | H | IW |
| 2.6 | 3.42 | 2.91 | PS | PS | PS | D | PS |
| 2.06 | 3.53 | 4 | PS | PS | PS | D | PS |

FIGURA 5.7: Captura de tablas de casas de apuestas con mejores cuotas

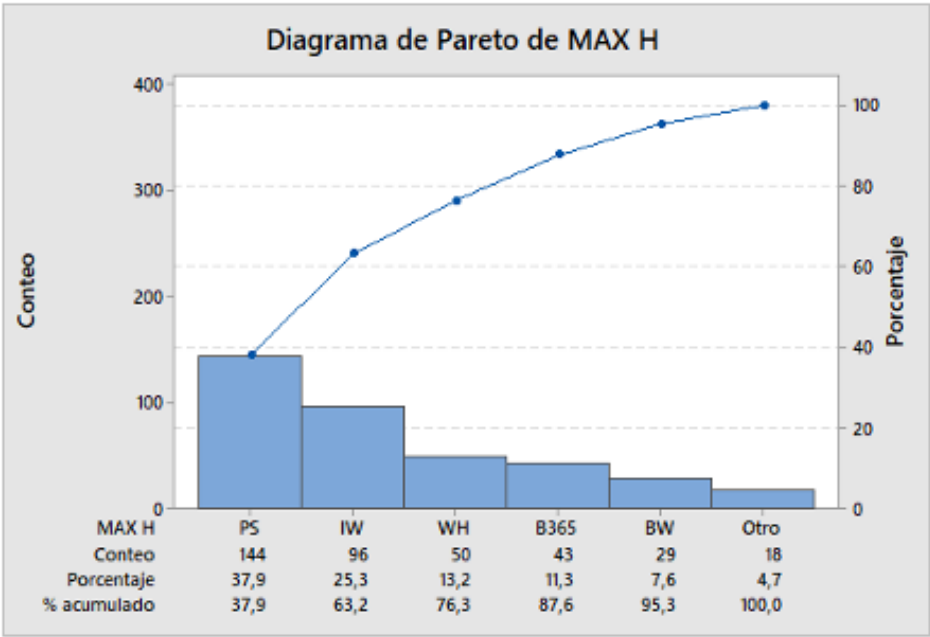


FIGURA 5.8: Casas de apuestas que mejor pagaban al local.

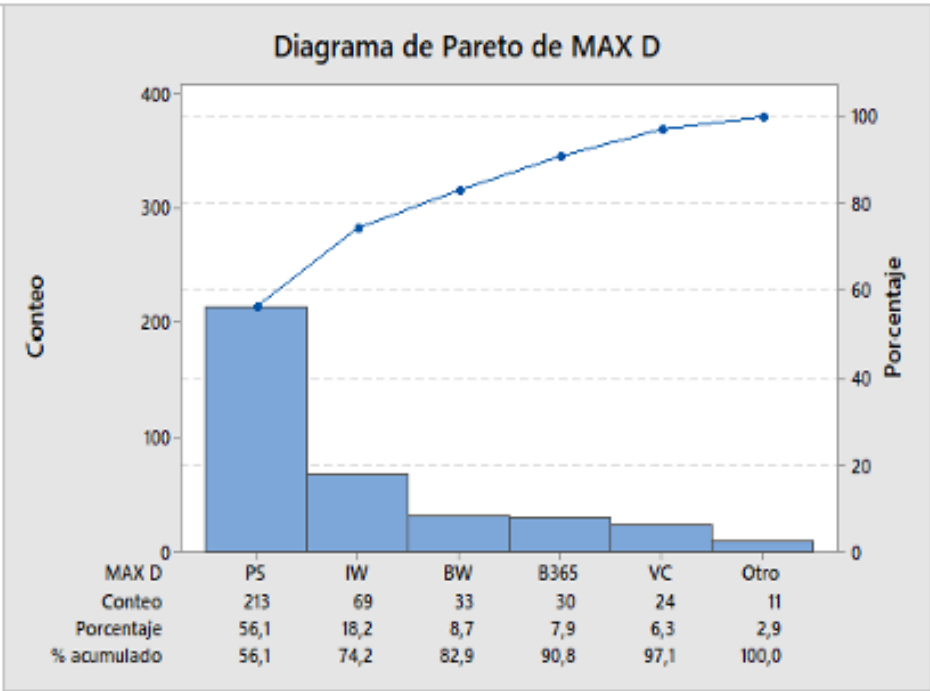


FIGURA 5.9: Casas de apuestas que mejor pagaban el empate.

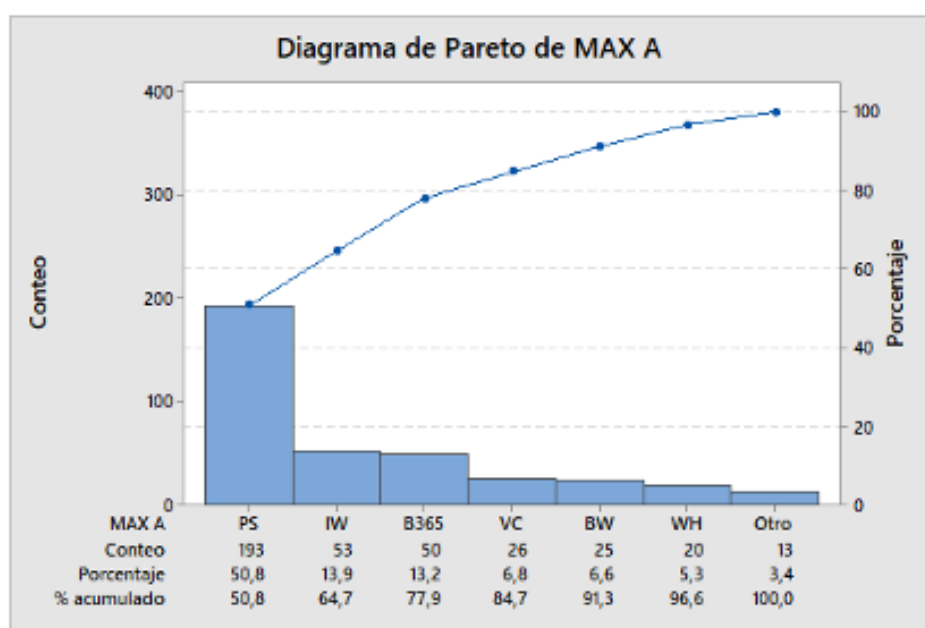


FIGURA 5.10: Casas de apuestas que mejor pagaban al visitante.

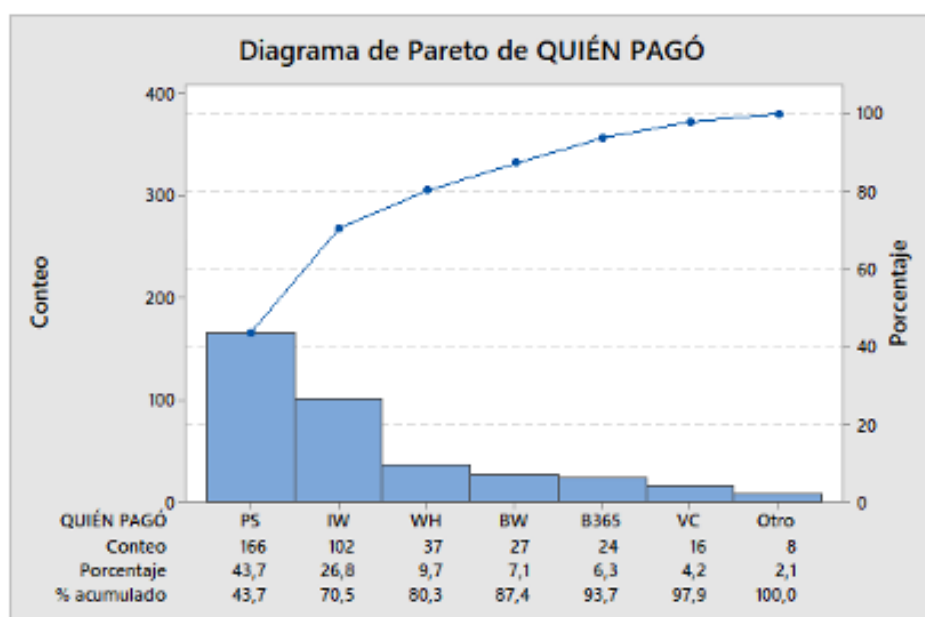


FIGURA 5.11: Casas de apuestas que ofrecían la mejor cuota del resultado que finalmente se dio

En todos los pasos las casas de apuestas Pinnacle (PS) e Interwetten (IT) ocupan las primeras dos posiciones. Aunque sean las casas con mejores cuotas, hay que insistir en que ambas siguen teniendo ventaja, ya que como se ha visto en la figura 5.4, por si solas y sin combinar con el resto de casas tienen rentabilidad inferior a la unidad.

Quizá si se retira a Pinnacle de la lista de casas de apuestas, de seguir habiendo SureBets está claro que o el número de encuentros que cumplan las condiciones será inferior, o la rentabilidad media será inferior, o incluso ambas a la vez. Podría darse el caso que Pinnacle disponga de unas cuotas suficientemente altas como para permitir SureBets, y que sin ellas no existirían.

Se prueba a repetir los mismos pasos sin la presencia de Pinnacle, en las figuras 5.12, 5.13 y 5.14 se muestran los resultados.

| MAX H SIN PS | MAX D SIN PS | MAX A SIN PS | Rentabilidad |
|--------------|--------------|--------------|--------------|
| 2,3 | 3,25 | 3,4 | 0,9647015696 |
| 2,75 | 3,25 | 2,8 | 0,9723181036 |
| 8,5 | 3,8 | 1,6 | 0,994233446 |
| 3,13 | 3,2 | 2,6 | 0,9836710604 |
| 1,2 | 7,5 | 17 | 0,9751435899 |
| 2,2 | 3,25 | 3,8 | 0,9752291789 |
| 6,5 | 4,6 | 1,53 | 0,9757715914 |
| 2 | 3,6 | 4,33 | 0,9913553811 |
| 2,55 | 3,4 | 2,88 | 0,9675858732 |
| 2,05 | 3,5 | 3,9 | 0,9709397726 |
| 1,08 | 13 | 36 | 0,970280314 |
| 6 | 3,6 | 1,8 | 1 |
| 2,5 | 3,2 | 3,4 | 0,9934235362 |
| 1,07 | 19 | 46 | 0,9911293919 |
| 1,7 | 4,6 | 6,5 | 1,042242071 |
| 2,3 | 3,4 | 3,5 | 0,9856003785 |
| 3,6 | 3,3 | 2,25 | 0,9753718605 |
| 2,3 | 3,3 | 3,5 | 0,9770109328 |
| 1,75 | 4,2 | 5,4 | 1,005318133 |
| 1,9 | 3,8 | 5 | 1,010642061 |
| 1,22 | 6,5 | 21 | 0,9792976477 |

FIGURA 5.12: Captura de tabla de rentabilidad sin la casa de apuestas Pinnacle.

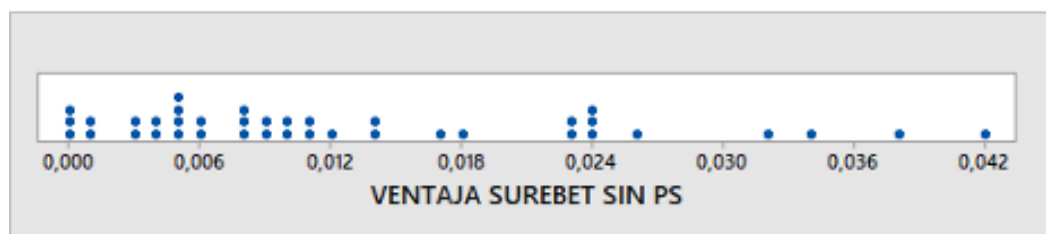


FIGURA 5.13: Boxplot de ventaja de surebet sin Pinnacle

| Estadísticas | | | | | | | | |
|------------------------|---------|-----|---------|----------------------------|-----------|---------|---------|---------|
| Variable | N | N* | Media | Error estándar de la media | Desv.Est. | Mínimo | Q1 | Mediana |
| VENTAJA SUREBET SIN PS | 39 | 341 | 0,01277 | 0,00179 | 0,01116 | 0,00006 | 0,00473 | 0,00932 |
| Variable | Q3 | | Máximo | | | | | |
| VENTAJA SUREBET SIN PS | 0,02303 | | 0,04224 | | | | | |

FIGURA 5.14: Captura de estadísticas descriptivas de la ventaja sin Pinnacle

El número de partidos con SureBet ha cambiado a 39. Hay menos, como era de esperar, pero sigue habiendo. Por lo tanto, ante una hipotética desaparición de la casa de apuestas Pinnacle, se podría seguir aplicando la estrategia.

- Una posible estrategia para aumentar la capacidad de ganancia, pero incorporando algo de riesgo, sería no fijando las ganancias iguales para cualquier resultado. Con la información del modelo predictivo y esta idea, quizá se podría llegar a una opción de mayor rentabilidad.

5.2.2. Martingala

El término martingala corresponde a una astucia mayormente conocida por su aplicación en la ruleta. La manera de operar es apostar una cantidad fija a un resultado que en futuros eventos mantenga la probabilidad de éxito. Por ejemplo, podría ser el caso de apostar a número impar. En caso de perder, se dobla la apuesta y se vuelve a apostar al mismo resultado, si se volviese a fallar, de nuevo se doblaría la apuesta anterior. El objetivo de doblar la apuesta en caso de fallo es recuperar las inversiones anteriores y la más reciente. Cuando se acierte, la secuencia queda cerrada, y si se desea se vuelve a comenzar. Se muestra un ejemplo en la tabla 5.10, donde sus unidades son monetarias.

| Inversión | Inversión acumulada | Éxito | Ganancia | Balance |
|-----------|---------------------|-------|----------|---------|
| 1 | 1 | NO | 0 | -1 |
| 2 | 3 | NO | 0 | -3 |
| 4 | 7 | NO | 0 | -7 |
| 8 | 15 | NO | 0 | -15 |
| 16 | 31 | NO | 0 | -31 |
| 32 | 63 | SÍ | 32·2=64 | +1 |

CUADRO 5.10: Ejemplo de aplicación de Martingala

Este método tiene dos principales inconvenientes:

- La banca no permite hacer apuestas sobrepasando un límite establecido, por lo que si tras varios intentos no hay éxito, no sería posible recuperar las pérdidas. La repetición de sucesos sin éxito es poco probable, pero las probabilidades no son nulas.
- Demasiado riesgo para balances tan pequeños. La cantidad a ganar, independientemente de cuántos intentos se lleven, será la primera cantidad apostada. Cuanto mayor sea la primera cantidad apostada, mayor probabilidad hay de superar el límite fijado por la banca y quebrar.

Para una situación en que la banca no permita cantidades superiores a los 500€:

Para una primera apuesta de 1€, en caso de obtener 8 resultados desfavorables el método habría quebrado, como se muestra en la ecuación 5.13 y su desarrollo. Las probabilidades de que esto suceda son las mostradas en la ecuación 5.14 y su desarrollo.

$$\text{Inversión acumulada en el octavo intento} = \sum_{n=1}^8 2^n = 479 \quad (5.13)$$

$$\text{Probabilidad de 8 intentos desfavorables consecutivos} = \left(\frac{18+1}{37}\right)^8 = 0,483\% \quad (5.14)$$

Martingala en el fútbol

En las apuestas de fútbol esta idea también se puede aplicar. Además se parte de una ventaja respecto a la ruleta, y es que hay muchas cuotas superiores que las de la ruleta, por lo que no sería necesario doblar la apuesta para recuperar la inversión. Será necesario analizar de nuevo algunos datos y ver dónde sería interesante apostar, y cómo serían los incrementos. Cabe decir que las que cuotas son superiores también tienen una probabilidad menor de éxito.

Dado que se quiere evitar llegar al tope establecido por la casa de apuestas, interesará un tipo de apuesta con probabilidad de éxito que no sea baja, y de variabilidad de cuota baja. Con las estadísticas descriptivas de los valores de las cuotas y de los resultados de partido, mostrados en las figuras 5.15 y 5.16 se intentará encontrar la mejor opción para aplicar la Martingala.

| Estadísticas | | | | | | | | | |
|---------------------|--------|----|-------|----------------------------|-----------|--------|-------|---------|-------|
| Variable | N | N* | Media | Error estándar de la media | Desv.Est. | Mínimo | Q1 | Mediana | Q3 |
| max cuota local | 380 | 0 | 3,151 | 0,164 | 3,187 | 1,050 | 1,672 | 2,150 | 3,000 |
| max cuota empate | 380 | 0 | 4,836 | 0,142 | 2,776 | 3,100 | 3,440 | 3,750 | 4,800 |
| max cuota visitante | 380 | 0 | 6,612 | 0,417 | 8,123 | 1,120 | 2,670 | 3,900 | 6,218 |
| Variable | Máximo | | | | | | | | |
| max cuota local | 26,000 | | | | | | | | |
| max cuota empate | 23,540 | | | | | | | | |
| max cuota visitante | 51,000 | | | | | | | | |

FIGURA 5.15: Captura de estadísticas descriptivas de las máximas cuotas para cada tipo de apuesta

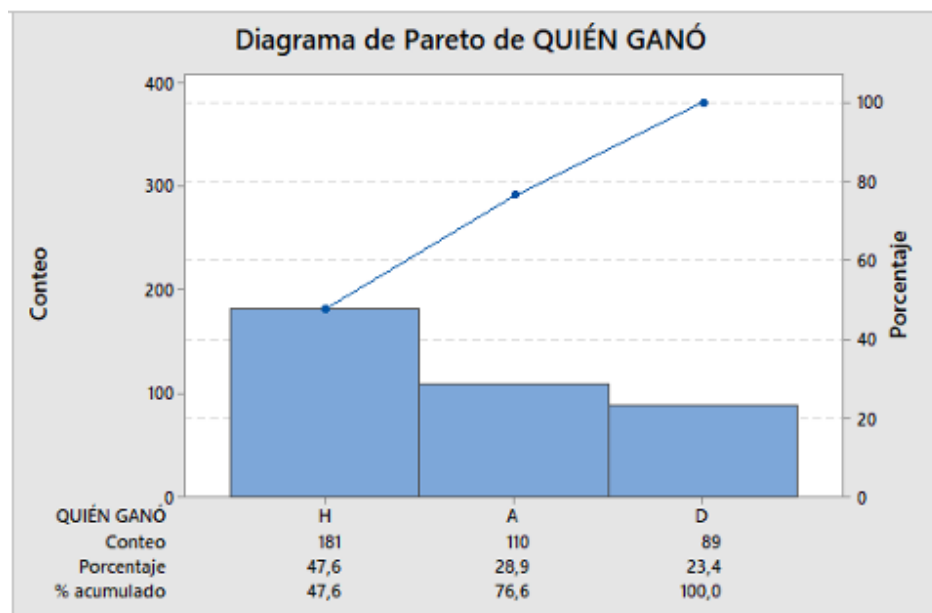


FIGURA 5.16: Diagrama de Pareto del resultado final, con intención de usarse para seleccionar qué martingala utilizar

Como era de esperar, las mayores cuotas las tiene el equipo visitante, pero también la desviación tipo. Hay más partidos que acaban en victoria visitante que en empate. Lo cual erróneamente puede hacer pensar que en una comparación de empate y visitante sería preferible la opción visitante.

Para saber la mejor opción se pueden describir las cuotas pagadas estratificadas según la opción de apuesta acertada, como se muestra en la figura 5.17.

| Estadísticas | | | | | | | | | | |
|----------------|-----|--------|----|-------|----------------------------|-----------|--------|-------|---------|-------|
| Variable | FTR | N | N* | Media | Error estándar de la media | Desv.Est. | Mínimo | Q1 | Mediana | Q3 |
| CUÁNTO SE PAGÓ | A | 110 | 0 | 3,133 | 0,335 | 3,512 | 1,120 | 1,722 | 2,675 | 3,518 |
| | D | 89 | 0 | 4,165 | 0,172 | 1,622 | 3,100 | 3,400 | 3,650 | 4,160 |
| | H | 181 | 0 | 2,105 | 0,109 | 1,471 | 1,050 | 1,390 | 1,780 | 2,240 |
| Variable | FTR | Máximo | | | | | | | | |
| CUÁNTO SE PAGÓ | A | 36,000 | | | | | | | | |
| | D | 13,400 | | | | | | | | |
| | H | 15,000 | | | | | | | | |

FIGURA 5.17: Captura de estadísticas descriptivas de la cuota pagada en los tipos 1, X y 2

Pruebas

A continuación se analiza el comportamiento y las cantidades a apostar para las condiciones:

1. La primera apuesta es de 0.5€
2. A partir de la segunda apuesta, incluida, el balance neto en caso de ganar se ha de ajustar a 2 veces el valor de la primera apuesta. Es decir, la ganancia posible siempre estará 1€ por encima de la inversión acumulada.
3. Se realizan 3 análisis diferentes: el de apuestas a local, el de a empate y el de a visitante.

- Prueba de apuesta a local:

| Cantidad apostada | Inversión acumulada | Posible ganancia | Mejor cuota local |
|-------------------|---------------------|------------------|-------------------|
| 0,5 | 0,5 | 1,005 | 2,01 |
| 2,142857143 | 2,642857143 | 3,642857143 | 1,7 |
| 36,42857143 | 39,07142857 | 40,07142857 | 1,1 |
| 24,73544974 | 63,80687831 | 64,80687831 | 2,62 |
| 88,77654563 | 152,5834239 | 153,5834239 | 1,73 |
| 667,7540171 | 820,337441 | 821,337441 | 1,23 |
| 198,8710511 | 1019,208492 | 1020,208492 | 5,13 |
| 300,0613212 | 1319,269813 | 1320,269813 | 4,4 |
| 2031,184328 | 3350,454142 | 3351,454142 | 1,65 |
| 4352,537846 | 7702,991988 | 7703,991988 | 1,77 |
| 5926,147683 | 13629,13967 | 13630,13967 | 2,3 |
| 9876,912805 | 23506,05248 | 23507,05248 | 2,38 |
| 13206,20926 | 36712,26173 | 36713,26173 | 2,78 |
| 4895,101564 | 41607,3633 | 41608,3633 | 8,5 |
| 18999,25265 | 60606,61594 | 60607,61594 | 3,19 |
| 275489,1634 | 336095,7793 | 336096,7793 | 1,22 |
| 280080,6494 | 616176,4287 | 616177,4287 | 2,2 |
| 112032,2598 | 728208,6885 | 728209,6885 | 6,5 |
| 728209,6885 | 1456418,377 | 1456419,377 | 2 |

FIGURA 5.18: Captura de tabla de las cuentas para una martingala al local

Debido a las bajas cuotas de los locales, en caso de no acertar relativamente pronto, las cantidades a apostar para recuperar las inversiones crecen muy rápidamente. Supone un riesgo preferiblemente evitable.

■ Prueba de apuesta al empate:

| Cantidad apostada | Inversión acumulada | Posible ganancia | Mejor cuota empate |
|-------------------|---------------------|------------------|--------------------|
| 0,5 | 0,5 | 1,7 | 3,4 |
| 0,5190311419 | 1,019031142 | 2,019031142 | 3,89 |
| 0,1682525952 | 1,187283737 | 2,187283737 | 13 |
| 0,9942198805 | 2,181503617 | 3,181503617 | 3,2 |
| 1,053478019 | 3,234981637 | 4,234981637 | 4,02 |
| 0,7699966612 | 4,004978298 | 5,004978298 | 6,5 |
| 1,679522919 | 5,684501216 | 6,684501216 | 3,98 |
| 2,631693392 | 8,316194609 | 9,316194609 | 3,54 |
| 3,105398203 | 11,42159281 | 12,42159281 | 4 |
| 4,140530937 | 15,56212375 | 16,56212375 | 4 |
| 7,077830662 | 22,63995441 | 23,63995441 | 3,34 |
| 10,41407683 | 33,05403124 | 34,05403124 | 3,27 |
| 15,135125 | 48,18915624 | 49,18915624 | 3,25 |
| 17,38132729 | 65,57048353 | 66,57048354 | 3,83 |
| 30,2593107 | 95,82979423 | 96,82979424 | 3,2 |
| 14,03330351 | 109,8630977 | 110,8630978 | 7,9 |
| 49,05446805 | 158,9175658 | 159,9175658 | 3,26 |
| 43,69332401 | 202,6108898 | 203,6108899 | 4,66 |
| 78,31188082 | 280,9227706 | 281,922771 | 3,6 |
| 116,4970132 | 397,4197839 | 398,4197852 | 3,42 |
| 157,4781774 | 554,8979612 | 555,8979662 | 3,53 |

FIGURA 5.19: Captura de tabla de las cuentas para una martingala al empate

Las cuotas del empate se podrían describir como suficientemente altas para que la inversión no se dispare y las cantidades a apostar estén dentro de los límites establecidos por las casas de apuestas.

- Prueba de apuesta al visitante:

| Cantidad apostada | Inversión acumulada | Posible ganancia | Mejor cuota visitante |
|-------------------|---------------------|------------------|-----------------------|
| 0,5 | 0,5 | 2,195 | 4,39 |
| 0,25 | 0,75 | 1,75 | 7 |
| 0,05833333333 | 0,8083333333 | 1,808333333 | 31 |
| 0,86523126 | 1,673564593 | 2,673564593 | 3,09 |
| 0,6076283167 | 2,28119291 | 3,28119291 | 5,4 |
| 0,1691336552 | 2,450326565 | 3,450326565 | 20,4 |
| 4,312908206 | 6,763234772 | 7,763234772 | 1,8 |
| 7,763234771 | 14,52646954 | 15,52646954 | 2 |
| 2,838477065 | 17,36494661 | 18,36494661 | 6,47 |
| 4,512271894 | 21,8772185 | 22,8772185 | 5,07 |
| 8,473043889 | 30,35026239 | 31,35026239 | 3,7 |
| 13,06260933 | 43,41287172 | 44,41287172 | 3,4 |
| 24,13743028 | 67,550302 | 68,550302 | 2,84 |
| 114,2505028 | 181,8008048 | 182,8008045 | 1,6 |
| 114,2505025 | 296,0513073 | 297,0513064 | 2,6 |
| 18,56570665 | 314,6170139 | 315,617013 | 17 |
| 104,5089445 | 419,1259584 | 420,1259569 | 4,02 |
| 778,0108538 | 1197,136812 | 1198,136715 | 1,54 |
| 334,6750466 | 1531,811859 | 1532,811713 | 4,58 |
| 802,5191391 | 2334,330998 | 2335,330695 | 2,91 |
| 778,4435034 | 3112,774501 | 3113,774014 | 4 |

FIGURA 5.20: Captura de tabla de las cuentas para una martingala al local al visitante

Las cuotas son variables y los aciertos poco frecuentes, aunque no se alcancen cantidades tan elevadas como en el caso de martingala a local, sigue siendo una apuesta arriesgada.

Debido a la seguridad que puede dar el no sobrepasar los límites de la banca, si se optase por aplicar alguna estrategia de Martingala, se recomendaría la de apuesta al empate.

5.3. Estrategias con exploración de datos: Caracterización de las casas de apuestas

En los datos que se utilizan y para el tipo de apuestas a tratar, se dispone de los valores de siete casas de apuestas diferentes. Todo lo que vendrá a continuación y se analizará en este apartado será un intento de aprovechar de toda esta información.

La primera idea que se pretende desarrollar es caracterizar las casas de apuestas. Como ya se ha ido comentando, los valores de las cuotas vienen definidos por el modelo predictivo

que cada casa tiene, además de las fluctuaciones debidas a los apostantes. Si se diera el caso de un partido en que varias casas de apuestas pagan una cantidad cercana a X por el local, y otras pagan una cantidad cercana a Y por el local, con una cierta diferencia entre X e Y , vendría a decir que según los modelos que utilizan las casas, discrepan en la probabilidad de darse este resultado.

Entonces, se pretende determinar qué casa de apuestas acierta en un porcentaje superior al resto y bajo qué condiciones. También se estudiará el valor esperado o esperanza matemática de la implementación de dicha estrategia.

En primer lugar se intentará encontrar características comunes en las cuotas para cada uno de los resultados posibles. En la base de datos se recogen datos del desarrollo del partido tales como por cada equipo la cantidad de saques de esquina, de disparos, de disparos a puerta, de goles al descanso y de goles al finalizar el partido. La intuición dice que habrá relación entre estos datos, aunque estaría por probar si es así. Aún así, no es esto lo que se pretende, sino relacionar características de antes del partido, con la del resultado final, y por ello se utilizarán las cuotas.

Se hará uso de una técnica estadística capaz de reducir la dimensionalidad de varias variables. Se conoce como análisis de componentes principales, y técnicamente no será explicada ahora, sino en el capítulo 6, donde sí será necesario entender mejor su significado.

Mediante un análisis multivariante se obtienen las componentes principales. Como variables iniciales se utilizan las 3 siguientes: valor medio de la cuota de local, valor medio de la cuota de visitante y valor medio de la cuota correspondiente a que no haya más de 2.5 goles durante el partido. Pese a ser dimensión 3 y pretender obtener 3 componentes principales, el resultado mostrará el mejor hiperplano, aquél que resuma mejor la información con sus 2 primeras componentes. En la figura 5.21 se muestra la distribución y la elongación de los ejes de las 3 variables, en el hiperplano construido con las dos primeras componentes principales. La primera explica un 56.5 % y la segunda un 38.8 %, en total un 95.3 % de la información.

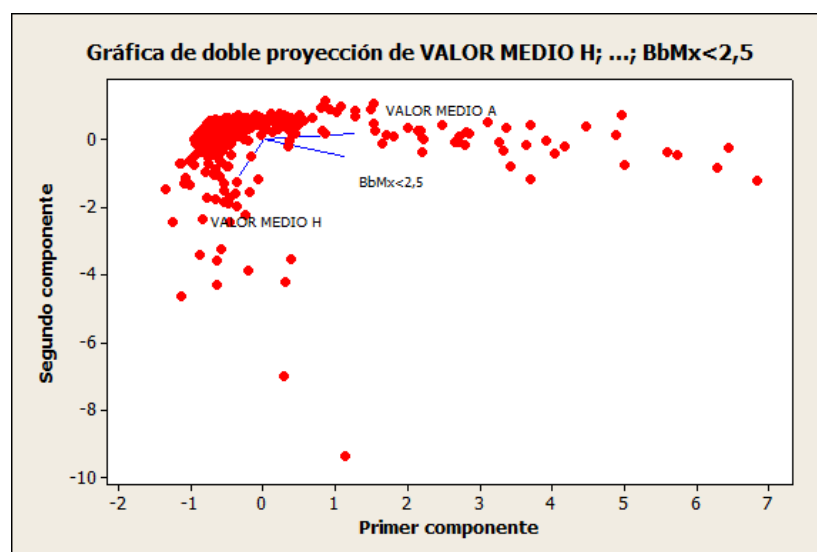


FIGURA 5.21: Dispersión de datos en plano PC2-PC1 y ejes de las variables

La primera componente parece diferenciar:

- $PC1 > 0$: condición de local favorito y partido con más de 2.5 goles
- $PC1 < 0$: apenas se podría decir que marca la condición de visitante favorito
- $PC2 > 0$: explica ligeramente la condición de local favorito
- $PC2 < 0$: explica la condición de visitante favorito, y ligeramente la de partido con más de 2.5 goles

Interesa graficar la nube de observaciones en el mismo gráfico PC2-PC1 diferenciando las observaciones según FTR (resultado final), como se muestra en la figura 5.22.

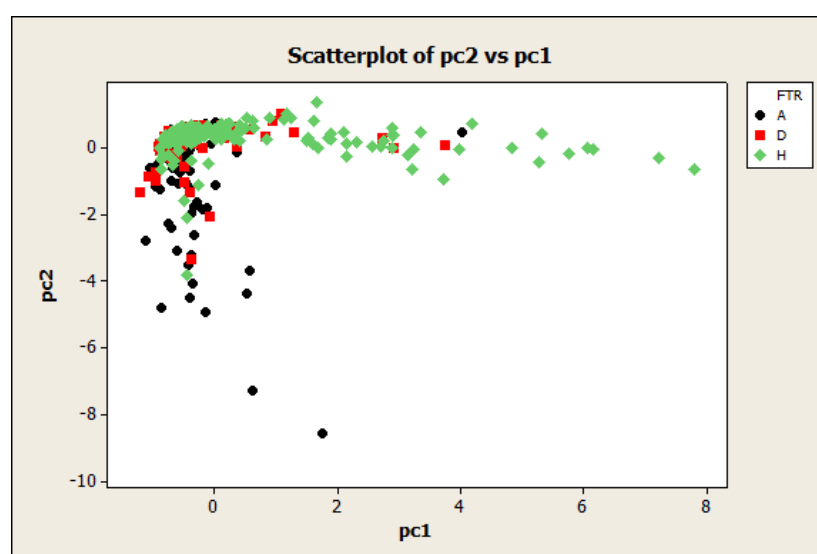


FIGURA 5.22: Diagrama con partidos pinados según FTR en el plano PC2-PC1

Los partidos donde terminó ganando el equipo local se encuentran con una PC2 nula y PC1 cercana a cero o superior. Según la información que dan las PC corresponde a encuentros con el local como favorito y donde se predecían más de 2.5 goles. Estos resultados son relativamente coherentes.

Los partidos en los que el resultado final fue victoria del visitante se encuentran con PC1 nula y PC2 negativa. Corresponde a encuentros donde el favorito era el visitante y además se predecían más de 2.5 goles. Los puntos más alejados del resto de victorias visitantes corresponden a encuentros de equipos punteros.

Los empates están situados cerca del origen. Corresponden a partidos donde ni local ni visitante eran claramente favoritos, y además no se predecían más de 2.5 goles. Esta información puede acercar a una buena estrategia para apostar al local, al empate o al visitante.

Inicialmente se hará un análisis exploratorio de empates, posteriormente de victorias visitantes y por último de victorias locales. Seguidamente se aplicaran las estrategias que parezcan convincentes en la temporada siguiente, la 2017-18. Aquellas que se mantengan se evaluarán por último en las dos temporadas previas a la de estudio, la 2014-15 y la 2015-16. Anticipar que se hicieron muchas pruebas para apostar a local, a visitante y a empate. Pero por el método de fijar estrategias los resultados entre temporadas no se mantenían, únicamente en algunas estrategias de empate. Dado que la intención es mostrar resultados interesantes y/o que muestren algún aprendizaje, no se procederá a añadir el seguido de pruebas que se realizó con local y visitante.

5.3.1. Empate

Las componentes principales no pudieron mostrar condiciones bastante concretas en las que se da el empate. Sabemos que mayormente son partidos en partidos parejos, y la pregunta es cuán de parejos, en qué rango nos interesa apostar.

En la figura 5.23 se muestran histogramas de las cuotas de todos los empates, y en la figura 5.24 de todos los empates estratificados según si el partido acabó en empate o no. El valor medio de las cuotas es de 4.495, aunque el que no interesa tanto como el valor medio cuando realmente se produce el empate, que es de 3.918. Observando la figura 5.24 se ve que el rango de las dos primeras columnas parece el mejor, por la menor discrepancia entre resultados con éxito y sin.

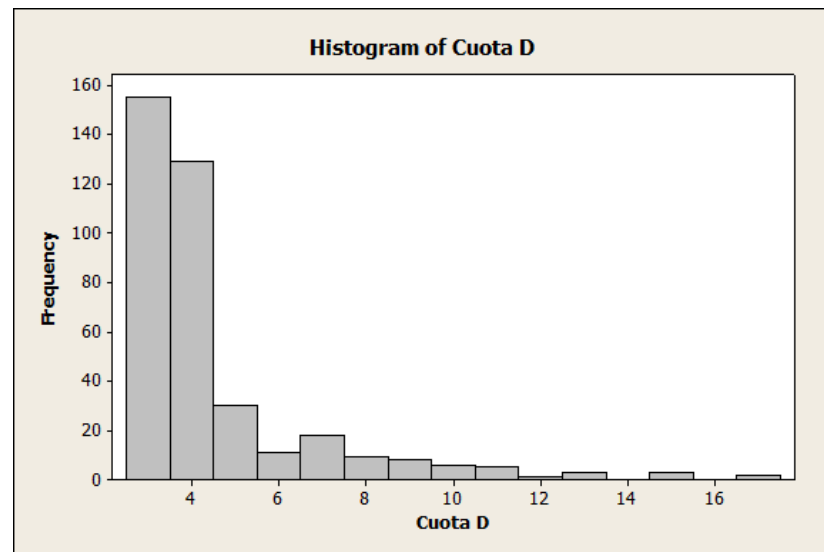


FIGURA 5.23: Histograma del valor de las cuotas del empate

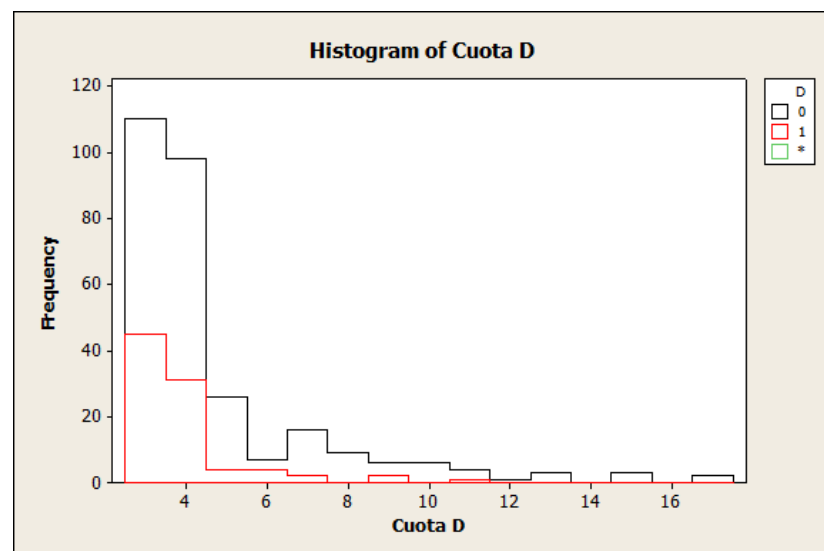


FIGURA 5.24: Histograma del valor de las cuotas del empate, diferenciando según si se dio empate o no

Sin embargo, aquí no se está considerando que debido al valor de las cuotas de empate, siempre superior a tres, un solo acierto puede compensar tres fracasos. Teniendo esto en cuenta faltaría ponderar los casos de éxito con su propio valor de cuota, y los de fracaso con ponderación unitaria. Aplicando esto se obtiene un nuevo histograma, con frecuencias y con grupos, mostrado en la figura 5.25. Parece que haya intervalos con beneficios, como la primera columna o el conjunto de las 3 primeras.

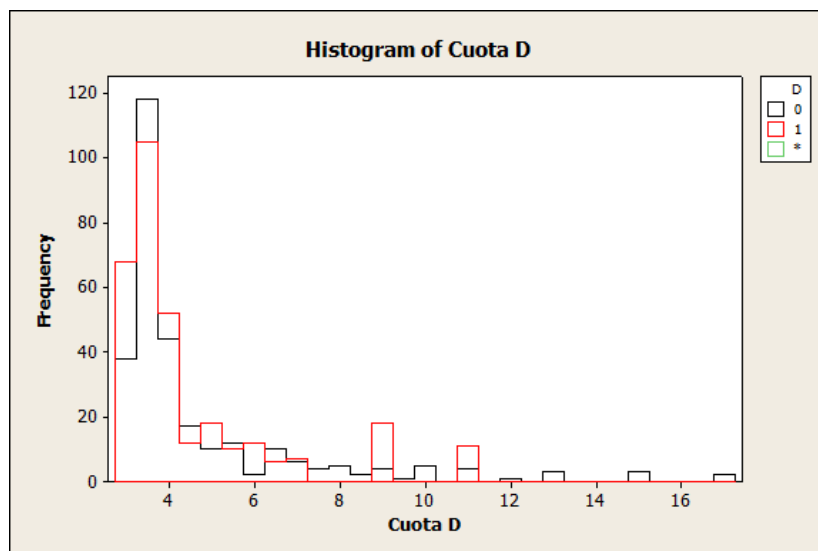


FIGURA 5.25: Histograma de los valores de las cuotas, con frecuencias y con separación según acierto

Este pequeño análisis está bien porque puede mostrar que es cierto que cuando los partidos son más parejos es cuando hay más empates, y también cuando apostar por ellos da más beneficios.

Un error que quizá se ha cometido hasta ahora es tratar de estudiar los empates mediante el valor de cuota de estos. En la tabla 5.11 se muestra los valores de las cuotas para dos partidos de la primera jornada.

| Equipo local | Equipo visitante | Cuota local | Cuota empate | Cuota visitante |
|--------------|------------------|-------------|--------------|-----------------|
| Deportivo | Eibar | 1.95 | 3.4 | 4.2 |
| Granada | Villarreal | 2.5 | 3.2 | 3 |

CUADRO 5.11: Ejemplo de cuotas. Importancia de la diferencia de cuotas de local y visitante.

En estos dos partidos las cuotas de los empates son bastante similares, y la información que estaríamos extrayendo sería similar. Sin embargo el motivo por el que tienen estas cuotas es muy diferente. En el primer encuentro hay un claro favorito y en el segundo no. Es por ello que no se va a trabajar con el valor en sí de la cuota de empate, sino con las diferencias que hay entre cuota de local y de visitante.

La manera más ergonómica de obtener resultados ha sido con la preparación de una hoja de Excel parametrizada según casa de apuestas y según diferencias de valores de cuotas. Es decir, una hoja que diera como opciones la selección de cualquier casa de apuestas (de la 1 a la 7) y de aquel valor máximo de diferencias de cuotas con el que se desea aplicar la

estrategia de apostar al empate. En la figura 5.26 se muestra un ejemplo, en la celda superior derecha se ha escogido aplicar la estrategia en partidos donde la diferencia de cuotas de la casa número 4 no difieran de más de 1.5 puntos. La casa número 4 se ha escogido en la celda superior central. En la hoja se configuró que la diferencia se estudiaba en valor absoluto. Más adelante también se mostrará la misma manera de operar pero únicamente apostando cuando el favorito sea el visitante. El gráfico muestra el acumulado de ganancias netas en caso de aplicar dicha estrategia. En este caso se observa que durante la primera mitad de la temporada resulta beneficiosa y acertada. En adelante la estrategia no resulta acertada y acaba con el equivalente a una esperanza matemática inferior a la unidad. Dado que los equipos a partir de mitad de temporada vuelve a enfrentarse a los mismos rivales que hasta el momento, podría ser que en ambas mitades hubiese comportamientos diferentes, y es algo que sobre todo con la observación de gráficos se tratará de determinar.

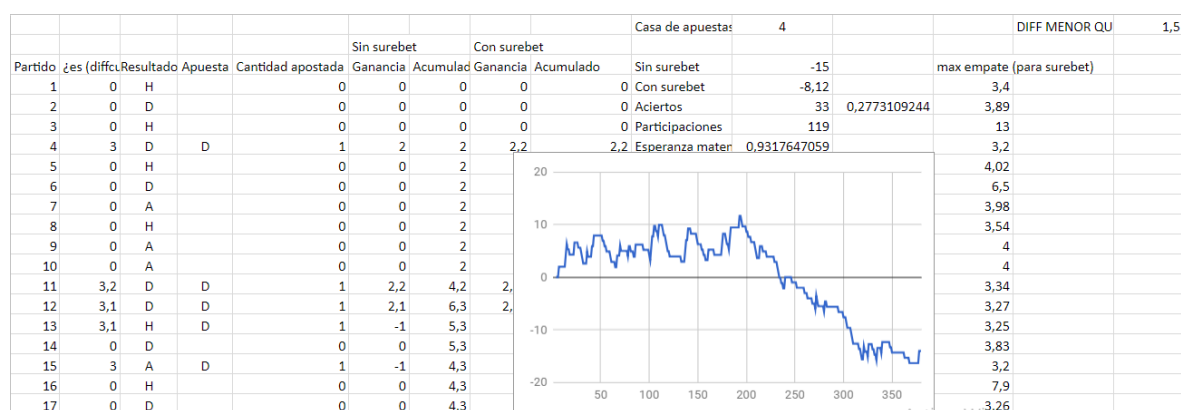


FIGURA 5.26: Captura de la operativa. Importancia de parametrizar.

En la tabla 5.12 se muestran los resultados obtenidos bajo la primera condición que se estudió, que fueron partidos bastante parejos según las cuotas. Concretamente diferencia de cuotas del equipo local y del visitante no podían diferir en más de un punto.

| <1 | Aciertos | Esp. Mat | Porcentaje aciertos | Ganancias (con cuota de la casa) | Ganancias (con) máxima cuota) | Particip. |
|-------------------|----------|-------------|------------------------|-------------------------------------|----------------------------------|-----------|
| 1 Bet365 | 24 | 1.13 | 33.80 | 6.1 | 9.55 | 71 |
| 2 Bet&Win | 24 | 1.07 | 32.00 | 2.5 | 5.55 | 75 |
| 3 Interwitten | 29 | 1.15 | 34.11 | 9.7 | 12.80 | 85 |
| 4 Ladbrokes | 29 | 1.13 | 33.78 | 4.7 | 9.92 | 74 |
| 5 Pinnacle | 23 | 1.10 | 32.85 | 6.6 | 7.52 | 70 |
| 6 William Hill | 26 | 1.07 | 32.26 | 2.6 | 6.02 | 81 |
| 7 BC Bet | 24 | 1.07 | 32.00 | 4.66 | 4.66 | 75 |

CUADRO 5.12: Resultados para la primera parametrización

Bajo esta condición y si en todas las temporadas fuera así, se podría afirmar que cualquier casa para apostar al empate es de fiar y sale rentable. En todos los encuentros la esperanza matemática es valor a la unidad, y la media de ganancias es de un 10.28 %.

Puntualizar que un añadido para conseguir más beneficios es que independientemente de con qué casa de apuestas se opte para saber si apostar o no, en caso de jugar se apostará en aquella que tenga la cuota más alta. Es por ello que hay dos columnas de ganancia, la segunda de ellas es con la operativa comentada. El valor esperado está calculado con ésta última.

Las estrategias probadas fueron varias y probadas bajo varios parámetros diferentes. Al referirse a parámetros se hace mención por ejemplo a las diferencias de valores entre cuotas de los dos equipos. Antes de mostrar resultados es necesario dejar claro que una de las intenciones de realizar tantas pruebas y quedarse con aquellas mejores, ha sido precisamente negar el buen hacer en estos pasos, ya que como se verá en el desarrollo de este punto, sobreajustar un modelo a unos datos puede resultar en que al probarlo con nuevos los resultados difieran del resultado deseado. Estas ideas están altamente relacionadas con el concepto “correlación no implica causalidad”.

En la temporada de estudio se vuelve a realizar lo mismo pero con variantes. Todas ellas son mostradas en la tabla 5.13.

- La primera de ellas es aumentar la diferencia entre cuotas aceptable hasta 1.5 puntos.

- La segunda es mantener la diferencia en 1 punto, pero únicamente evaluando aquellos partidos donde la cuota superior es la del local, es decir, que en un inicio las probabilidades se decantan hacia la victoria del visitante.
- Posterior a la segunda estrategia, se fue variando el parámetro de la diferencia, hasta llegar a una tercera situación, en que la diferencia no puede sobrepasar 0.6 puntos.
- Por último, se añade la condición necesaria de que el valor de la cuota de más de 2.5 goles esté por debajo de 2.1. Esto se traduce en que además de partidos parejos, se está buscando partidos donde se intuyen pocos goles. Esta acción se toma ya que las componentes principales marcaban que los empates suelen ir acompañados de partidos con pocos goles. El valor de 2.1 simplemente es aquél con el que se obtenían mejores resultados.

En cada casilla, perteneciente a una estrategia y una casa de apuestas, se encuentran 2 celdas. La superior indica cuál ha sido la esperanza matemática encontrada, y que interesa que sea superior a la unidad. La inferior hace referencia al porcentaje de acierto.

| Esp. Matemática | <1.5 | <1 | <0,6 | 1 + BbMax2.5 >2.1 |
|---------------------------|-------------------|--------------|----------------|--------------------------------|
| Porcentaje acierto | | | | |
| 1 Bet365 | 0.91 | 1.17 | 1.39 | 1.16 |
| | 27.59 | 34.88 | 41.32 | 34.92 |
| 2 Bet&Win | 0.96 | 1.09 | 1.27 | 1.04 |
| | 28.76 | 32.65 | 37.83 | 32.83 |
| 3 Interwitten | 0.94 | 1.20 | 1.05 | 1.18 |
| | 28.22 | 25.55 | 31.03 | 35.61 |
| 4 Ladbrokes | 0.93 | 1.21 | 1.198 | 1.15 |
| | 27.33 | 36.17 | 35.48 | 34.92 |
| 5 Pinnacle | 0.96 | 1.22 | 1.352 | 1.12 |
| | 28.97 | 36.36 | 40.00 | 33.89 |
| 6 William Hill | 0.97 | 1.07 | 1.621 | 1.10 |
| | 28.92 | 32.01 | 34.37 | 33.3 |
| 7 BC Bet | 0.96 | 1.14 | 1.208 | 1.08 |
| | 28.82 | 34 | 38.23 | 32.81 |

CUADRO 5.13: Resultados de varias estrategias con mejores resultados sobre la propia temporada

Primeras conclusiones

Posterior a estos resultados, varios de ellos interesantes, había que probarlos en la temporada siguiente. La realidad es que ambos indicadores disminuyeron bastante por lo general,

aún así en muchas ocasiones seguían siendo interesantes. Hay que recordar que se parte de una situación inicial de apostar al empate de un 23.4 % de acierto y de un valor esperado de 0.92. En cuanto a caracterización de la casa de apuestas, debido a que entre la temporada de pruebas y la de estudio hay discrepancias y poca estabilidad de los valores, no se va a considerar aún ninguna caracterización. La estrategia de apostar cuando las diferencias sean menores de 0.6 puntos sí que fue satisfactoria en la temporada de pruebas, es por ello que en el tabla 5.14 se muestran los indicadores para las últimas 4 temporadas. De nuevo, los resultados vuelven a ser satisfactorios, y al ya haber probado con más de 1000 partidos se puede establecer la primera estrategia de esta operativa. Dado que Bet365 es aquella que ofrece mejores resultados, se establece que será a esta casa de apuestas con la que habrá que comparar el valor de las cuotas.

En el gráfico 5.27 se muestra el acumulado de la estrategia a aplicar de las últimas 4 temporadas. Tanto los resultados como el transcurso de cada temporada son positivos.

| <0,6 | 2014-15 | 2015-16 | 2016-17 | 2017-18 (hasta jornada 32) | 2017-18 (temporada finalizada) |
|----------------|---------|---------|---------|-------------------------------|-----------------------------------|
| 1 Bet365 | 1.25 | 1.37 | 1.39 | 1.58 | 2.09 |
| | 38.09 | 41.67 | 41.32 | 47.36 | 47.36 |
| 2 Bet&Win | 1.09 | 1.31 | 1.27 | 1.36 | 1.76 |
| | 33.3 | 39.47 | 37.83 | 40.74 | 40.74 |
| 3 Interwitten | 1.3 | 1.18 | 1.05 | 1.34 | 1.68 |
| | 38.89 | 35.71 | 31.03 | 40.62 | 40.62 |
| 4 Ladbrokes | 1.46 | 1.54 | 1.20 | 1.11 | 1.38 |
| | 44.44 | 46.42 | 35.48 | 33.34 | 33.33 |
| 5 Pinnacle | 1.05 | 1.56 | 1.35 | 1.18 | 1,40 |
| | 31.57 | 46.67 | 40.00 | 35.29 | 35.29 |
| 6 William Hill | 1.16 | 1.48 | 1.62 | 1.16 | 1.53 |
| | 35.29 | 44.44 | 34.37 | 34.61 | 34.61 |
| 7 BC Bet | - | 1.61 | 1.21 | 1.07 | 1.36 |
| | - | 48.14 | 38.23 | 31.81 | 31.81 |

CUADRO 5.14: Resultados de la estrategia de caracterización escogida, aplicación en las últimas cuatro temporadas

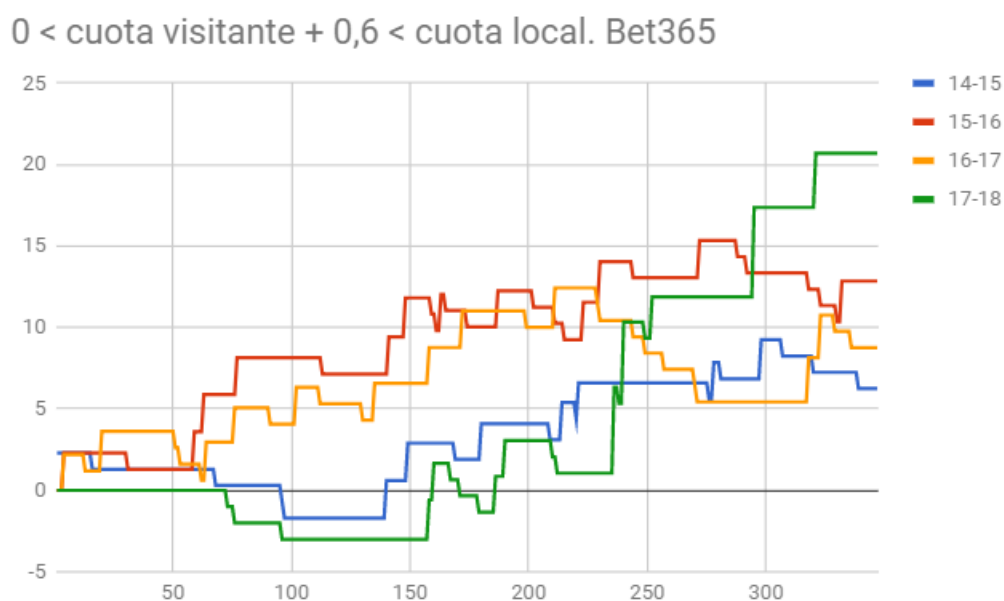


FIGURA 5.27: Evolución de los beneficios de la estrategia de caracterización conseguida, aplicación sobre las últimas 4 temporadas

Por si estos pasos pueden parecer excesivamente artificiales, se va a proceder a explicar el significado en términos cotidianos. Inicialmente se ha visto que cuando las cuotas de local y visitante no eran altas era cuando más se producían empates. Este hecho tenía relación con diferencias de cuotas bajas. Fuera de lo realizado en estos pasos, se entiende que cada casa de apuestas realmente es muy informativa de cómo concluirá un partido. Es por ello que si alguna tuviera modelos óptimos, ayudaría al usuario que quiera beneficiarse en la creación de una estrategia, en poder conseguirla. Evaluando varias estrategias que tenían sentido y con varias casas de apuestas, se llega a la conclusión que en un cierto intervalo, se puede hacer caso a lo que Bet365 en relación a la probabilidad de éxito del empate.

Segundas conclusiones

Correlación sin causalidad: A menos que se establezca la causalidad que reside detrás de la correlación, no se sabe qué factores pueden explicar dicha correlación, y por lo tanto la correlación sin causalidad es irrelevante.

Existe un ejemplo que presenta cómo se intentó estimar la esperanza de vida de los países europeos a partir de la cantidad de vino que consumían. Los resultados mostraban como aquellos países donde se consumía más vino tenían esperanzas de vida superiores. El factor determinante por supuesto que no era el vino, sino otros factores que sí tenían correlación, como la alimentación, la calidad de vida o el factor climatológico, entre otros.

De la misma manera aquí, cuando se establece que si las cuotas difieren de un cierto valor,

hay que preguntarse si ello tiene sentido. Dado que las estrategias han partido del significado extraído con las componentes principales, sí que había relación entre valor de cuotas y resultado. Sin embargo, en el proceso de buscar aquello que con estos datos se obtenga los mejores resultados, ha implicado llegar a estrategias sobreajustadas y que al testearlas con nuevos datos caían en porcentaje de aciertos y en esperanza matemática.

Capítulo 6

Diseño de estrategias a partir de técnicas de análisis de datos

Previo a las explicaciones teóricas de las técnicas utilizadas en este capítulo 6, en la tabla 6.1 se muestra una clasificación sobre las mismas técnicas a utilizar. Dicho cuadro puede servir para entender diferencias entre técnicas y/o anticipar lo que viene a continuación.

| Método | Uso de | Variables predictivas | Variables respuesta |
|-----------------------------|---------------------|-----------------------|---------------------|
| k-Means | Agrupación | PC1 PC2 PC3 | No usa |
| Discriminant Analysis | Agrupación | PC1 PC2 PC3 | FTR |
| Nominal Logistic Regression | Modelo de regresión | PC1 | FTR |

CUADRO 6.1: Resumen de las técnicas que se emplearán

En este apartado se presentarán tres técnicas de análisis de datos, y las variables predictivas ya no serán las cuotas, sino sus inversos, entendiendo que estos corresponden a la probabilidad de éxito de cada posible resultado. Cada una de las técnicas se caracterizará por el uso simultáneo de todas las variables predictivas. Dado que se trabaja con una alta dimensionalidad, se recurre a una técnica de reducción de la misma sin pérdida de información. Se está hablando del análisis de componentes principales.

6.1. Análisis de componentes principales

Cuando se recoge la información de una muestra de datos, lo más frecuente es tomar el mayor número posible de variables. Sin embargo, si se toman demasiadas sobre el conjunto

de observaciones la matriz de coeficientes de correlación será de una dimensión excesiva. En este caso es difícil visualizar relaciones entre las variables. Otro problema es que estas variables estén relacionadas o midan lo mismo desde diferentes puntos de vista. Un ejemplo sería en estudios médicos, la presión sanguínea a la salida del corazón y a la salida de los pulmones.

Entonces, APC se utiliza para conformar un número menor de variables no correlacionadas a partir del conjunto grande. La meta del análisis consiste en explicar la cantidad máxima de varianza con el menor número de componentes principales [2]. Las nuevas variables son combinación lineal de las anteriores y se van construyendo según el orden de importancia en cuanto a la variabilidad total que recogen de la muestra.

Para el caso, se realizará una demostración únicamente intuitiva:

Según la relación entre variables se opera de diferentes formas:

- Correlaciones: Se operará con matriz de correlaciones en caso de que el orden de magnitud de las variables no sea el mismo y los valores medios no sean similares
- Covarianzas: Se trabaja con matriz de covarianzas si las campanas de las variables son similares

Como ya se vió en el capítulo 4 hay significanzas entre variables de cuota local, de empate y de visitante, por lo que se procede a operar con matriz de correlaciones.

Con una matriz F_i^j de i filas, observaciones, y j columnas, variables, se determina cada elemento de la matriz de correlación muestral, tal y como queda expresado en la ecuación 6.3. Esta matriz, ya simétrica y por lo tanto diagonalizable, tiene unos valores propios, que son los mostrados en la ecuación 6.3.

$$R = [r_{ij}] \in M_{m \times m} \quad (6.1)$$

$$r_{ij} = \frac{\text{cov}(F_i, F_j)}{\sqrt{\text{var}(F_i)\text{var}(F_j)}} \quad (6.2)$$

$$\sum_{i=1}^m \lambda_i = 1 \quad (6.3)$$

A partir de aquí, los m valores propios recogen la cantidad de covarianza capturada, y también se les conoce como *scores*, y los vectores propios, también llamados *loadings* marcarán las direcciones ortogonales de las nuevas variables, las componentes principales.

En la figura 6.1 se muestra un plano sobre el que hay dos vectores que indican las direcciones principales de inercia, es decir, las direcciones de las componentes principales.

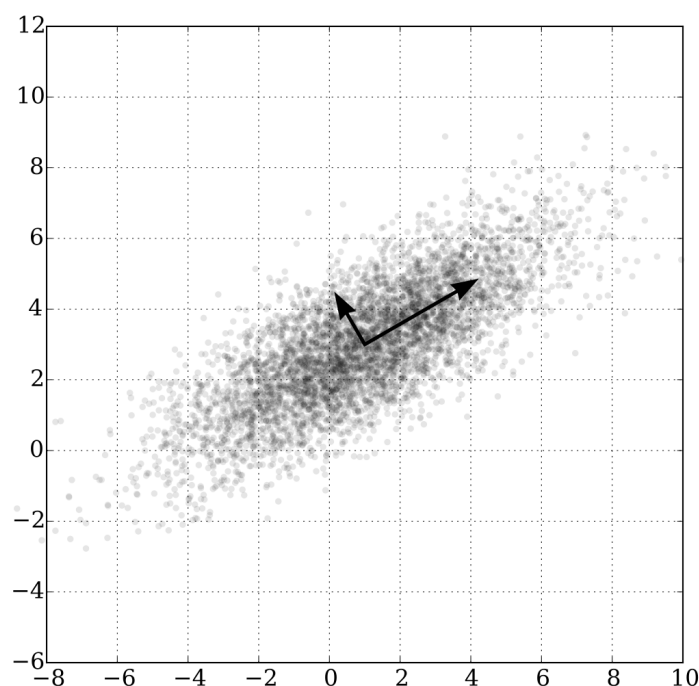


FIGURA 6.1: Ejemplo de las dos direcciones principales de inercia para una dispersión de puntos. Fuente: [3]

En resumen, ACP busca la proyección según la cual los datos queden mejor representados en términos de mínimos cuadrados. Esta convierte un conjunto de observaciones de variables posiblemente correlacionadas en un conjunto de valores de variables sin correlación lineal llamadas componentes principales. Los componentes se ordenan por la cantidad de varianza original que describen, por lo que la técnica es útil para reducir la dimensionalidad de un conjunto de datos. [4]

En el caso del estudio, aplicando la técnica a las 21 variables resulta haber una proyección de los datos, en un plano formado por los ejes de una primera y una segunda componente, que recoge el 99.5 % de la información, como se puede ver en la figura 6.2. En la figura 6.3 se muestran todas las observaciones representadas en dicha proyección. Por la forma que toma el gráfico se interpreta que los datos son composicionales. Los encuentros situados hacia la izquierda corresponden a partidos donde las probabilidades marcan una tendencia a local vencedor. En la derecha están aquellos donde es el visitante el que según probabilidades será vencedor. En ambos casos, cuanto más escorados estén, mayor serán las condiciones mencionadas. Se podría decir que la variable PC1 toma valores positivos cuando el favorito es el visitante, y valores negativos cuando lo es el local.

Principal Component Analysis: B365H; B365D; B365A; BWH; BWD; BWA; IWH; IWD; IWA

Eigenanalysis of the Correlation Matrix

376 cases used, 4 cases contain missing values

| | | | | | | | | | |
|------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| Eigenvalue | 14,997 | 5,895 | 0,034 | 0,018 | 0,015 | 0,011 | 0,010 | 0,004 | 0,003 |
| Proportion | 0,714 | 0,281 | 0,002 | 0,001 | 0,001 | 0,001 | 0,000 | 0,000 | 0,000 |
| Cumulative | 0,714 | 0,995 | 0,997 | 0,997 | 0,998 | 0,999 | 0,999 | 0,999 | 0,999 |

FIGURA 6.2: Captura de la salida de Minitab del análisis de componentes principales.

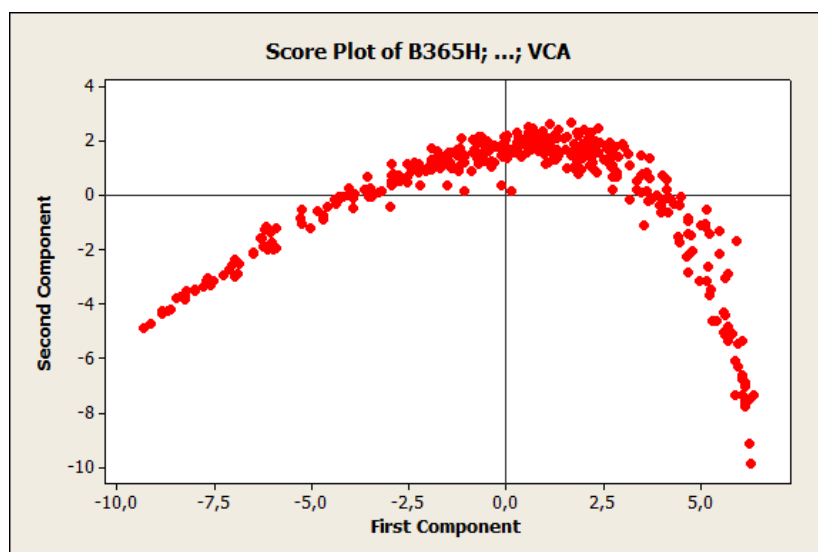


FIGURA 6.3: Dispersión de las probabilidades en el plano PC2-PC1

En los siguientes apartados se presentarán las técnicas. Para cada una de ellas se diferenciará una explicación formal y con mayor carácter matemático, una explicación de menor nivel y enlazando con nuestro caso, junto con la aplicación de la técnica, seguidamente los resultados sobre los mismos datos y los resultados sobre otros datos diferentes, y por último unas conclusiones.

6.2. Cluster k-Means

Esta técnica se emplea cuando los grupos se desconocen inicialmente. El procedimiento utiliza conglomerados no jerárquicos de observaciones según el algoritmo de MacQueen. Los conglomerados de K-medias funcionan mejor cuando hay suficiente información disponible para hacer buenas designaciones del conglomerado inicial.

El primer paso del algoritmo es la inicialización. Para ello se seleccionan k individuos de la población al azar. Esos individuos son los representantes de cada clúster y se llaman también centroides. Una vez hecho esto, el algoritmo repite un número de veces dos sencillos

pasos: Se evalúa cada observación, moviéndola hacia el conglomerado más cercano. El conglomerado más cercano es el conglomerado que tiene la distancia Eucladiana más pequeña entre la observación y el centroide del conglomerado.

Cuando un conglomerado cambia, al perder o ganar una observación, se vuelve a calcular el centroide del conglomerado. Este proceso se repite hasta que no se pueda mover más observaciones a un conglomerado diferente. En este punto, todas las observaciones se encuentran en su conglomerado más cercano de acuerdo al criterio anteriormente mencionado.

Los pasos se repiten hasta que los grupos se mantienen iguales o, lo que es lo mismo, no hay nuevas reasignaciones de individuos en el paso 1, y por lo tanto el algoritmo ha convergido.

Es un método utilizado en minería de datos. El problema es computacionalmente difícil (NP-hard). Sin embargo, hay eficientes heurísticas que se emplean comúnmente y convergen rápidamente a un óptimo local.

En la figura 6.4 se interpreta con sencillez el significado del centroide, y en la figura 6.5 se observan las iteraciones que emplea el método.

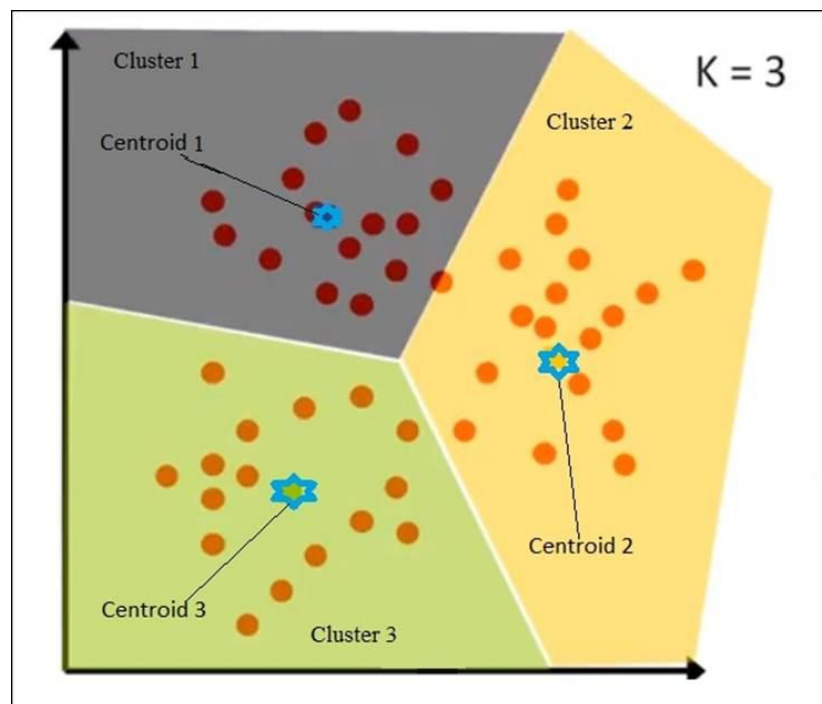


FIGURA 6.4: Muestra de centroides de grupos. Fuente: [5]

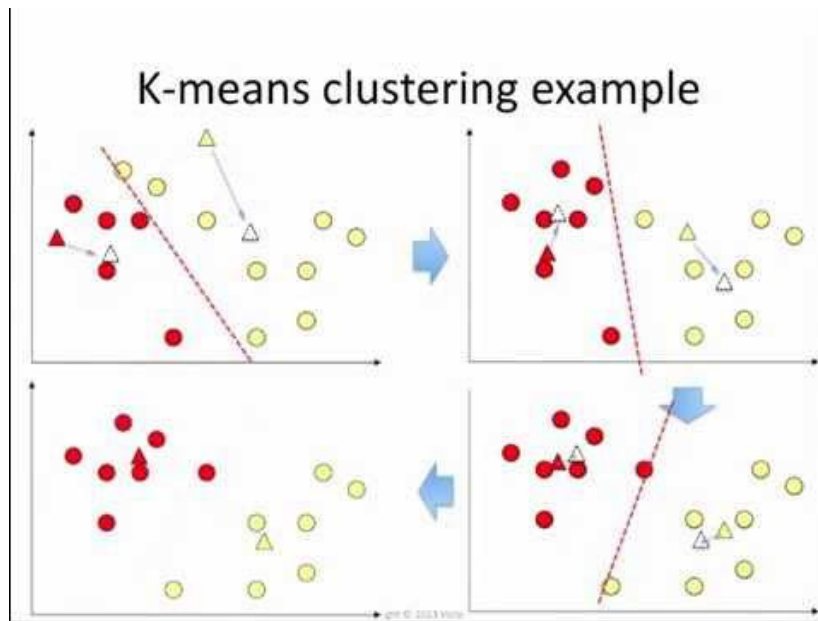


FIGURA 6.5: Iteraciones en búsqueda de convergir. Fuente [6]

6.2.1. Resultados

En el caso, se procede a hacer 3 y 4 grupos, en modelos diferentes, y evaluarlos. Al hacer 3 se vio que cada uno directamente representaba aquel resultado más probable. Es decir, un grupo agrupaba encuentros de local favorito, otro grupo lo agrupaba los de visitante favorito, y el restante eran encuentros parejos. Al implementar la técnica para que hiciese 4 grupos, éstos agrupaban situaciones diferentes. Tal y como se puede ver en la figura 6.6, y recordando el significado que en este ejercicio guarda PC1, se definen los grupos 3 y 4 como grupos identificadores de local favorito y visitante favorito, respectivamente. Los grupos 1 y 2 por su ubicación corresponden a partidos más parejos que los anteriores, y entre ellos se diferencian por tener una cierta decantación hacia el visitante, y hacia el local, respectivamente.

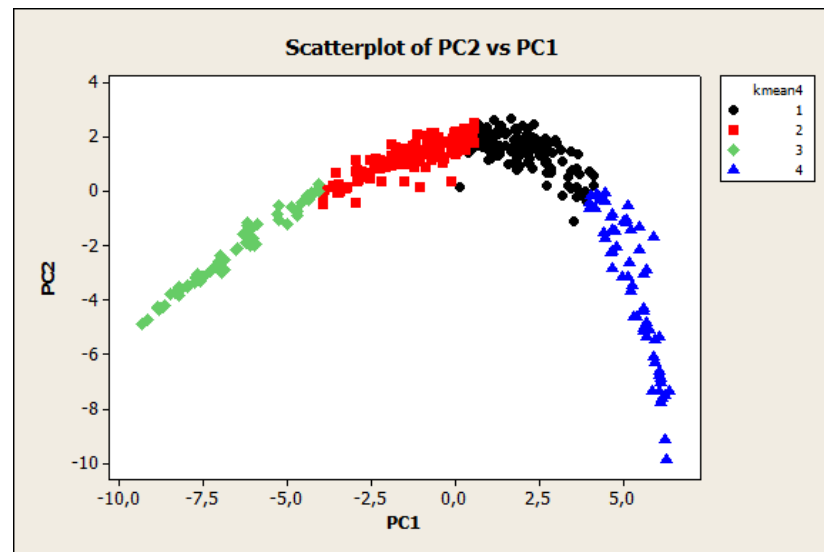


FIGURA 6.6: Grupos creados pintados. Plano PC1-PC2.

Hasta el momento se dispone de una técnica capaz de agrupar encuentros en 4 clasificaciones diferentes. En la figura 6.7, las columnas son los grupos y las filas son el resultado final. De los 52 encuentros que pertenecen al grupo 3, 43 de ellos sí finalizaron en victoria local, y de los 52 encuentros del grupo 4, 36 también acertados como victoria visitante. En el grupo 1, que en un inicio se había caracterizado por ser parejo, los resultados son muy dispersos, y en el grupo 2, hay una tendencia hacia la victoria local.

| Rows: FTR | | Columns: kmean4 | | | | |
|-----------|--|-----------------|-----|----|----|-----|
| | | 1 | 2 | 3 | 4 | All |
| A | | 58 | 14 | 1 | 36 | 109 |
| D | | 47 | 25 | 8 | 9 | 89 |
| H | | 55 | 73 | 43 | 7 | 178 |
| All | | 160 | 112 | 52 | 52 | 376 |

FIGURA 6.7: Captura de la tabla de correspondencia que relaciona grupos y resultados

En los gráficos 6.8 hasta 6.11 se muestran las evoluciones de la aplicación de los grupos 1, 2, 3 y 4 a cada una de las apuestas básicas posible, como son H, D y A. A partir de los gráficos se pretende:

- Entender comportamiento.
- Seleccionar aquellas estrategias con resultados positivos al final de la serie.
- Valorar la seguridad del punto anterior, a partir de la tendencia que tuvo.

1H, 1D y 1A

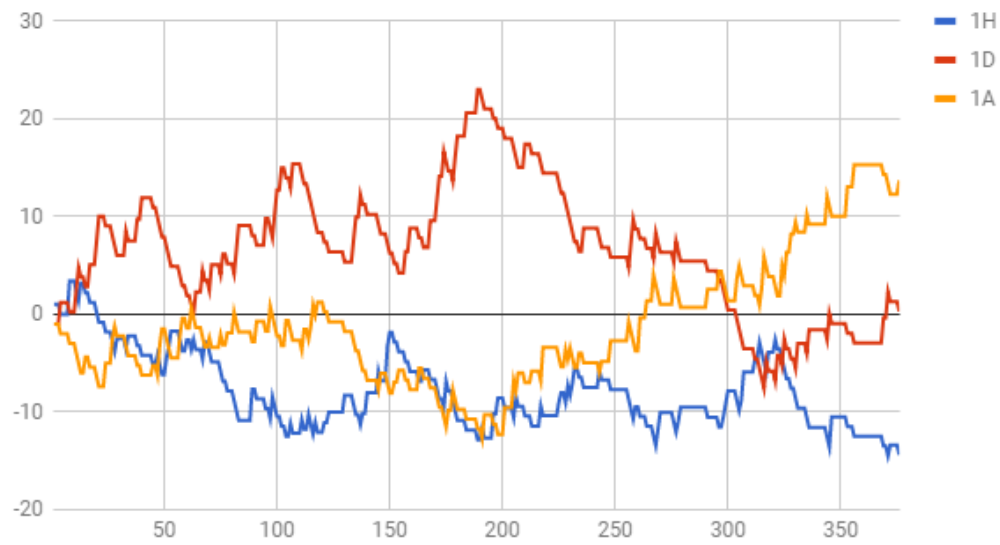


FIGURA 6.8: Evolución de beneficios acumulados apostando únicamente al grupo 1

2H, 2D y 2A

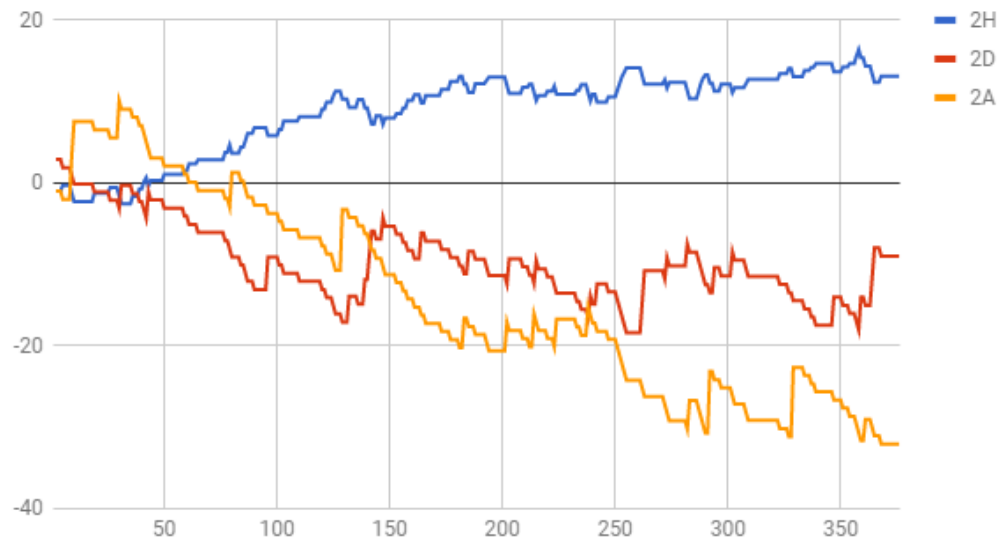


FIGURA 6.9: Evolución de beneficios acumulados apostando únicamente al grupo 2

3H, 3D y 3A

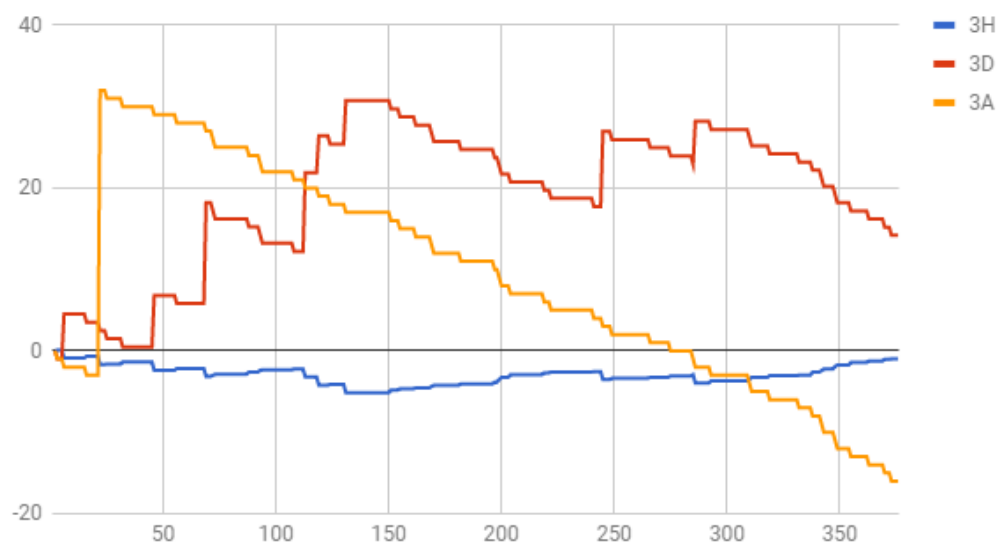


FIGURA 6.10: Evolución de beneficios acumulados apostando únicamente al grupo 3

4H, 4D y 4A

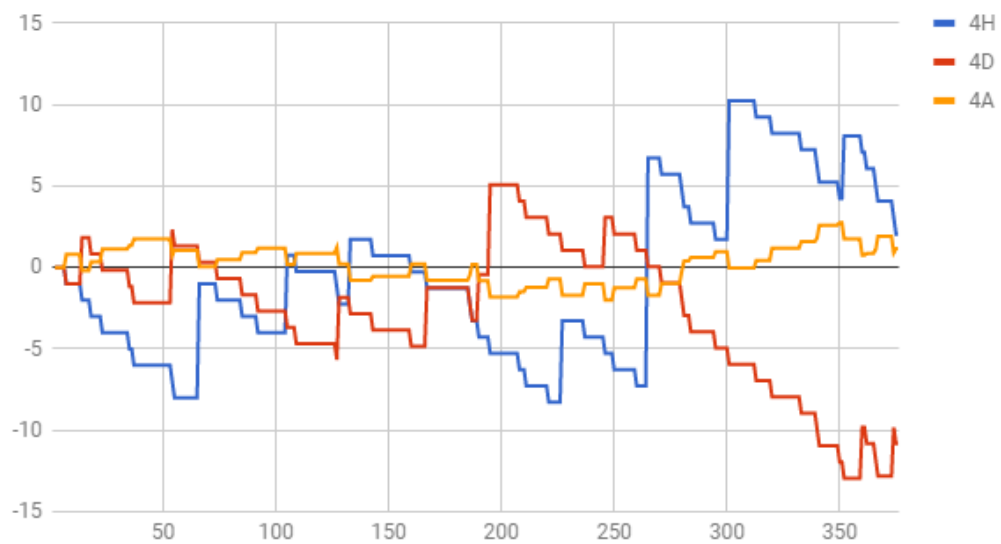


FIGURA 6.11: Evolución de beneficios acumulados apostando únicamente al grupo 4

De cara a ponerlo en práctica, la idea no es acertar en una mayor cantidad de ocasiones, sino establecer unas estrategias a largo plazo rentables, o al menos mejores que las de los casos base. Los beneficios finales de las pruebas para cada estrategia básica con cada grupo, que se han mostrado en las figuras 6.8, 6.9, 6.10 y 6.11 se muestran en la se resumen en la

tabla 6.2. En la tabla 6.3 se muestran aquellas estrategias que han tenido buenos resultados, y que se probarán en la temporada siguiente. Como ya se explicó en el capítulo 3, en 4.3, y a modo de recordatorio, la esperanza matemática o valor esperado resulta de dividir la suma recuperada entre la suma invertida, que al invertir 1€ por apuesta, coincide con el número de ocasiones invertidas.

| Beneficio | 1 | 2 | 3 |
|------------------|----------|----------|----------|
| H | -14.39 | 13.12 | -0.97 |
| D | 0.28 | -8.98 | 14.18 |
| A | 13.7 | -32.05 | -16 |

CUADRO 6.2: Beneficios obtenidos según grupos y apuesta a realizar, para k-means

| Definición de grupo | Apuesta a realizar | Beneficio | Esperanza matemática |
|----------------------------|---------------------------|------------------|-----------------------------|
| Parejo-visitante | Visitante | 13.7 | 1.085 |
| Parejo-local | Local | 13.12 | 1.117 |
| Local | Empate | 14.18 | 1.273 |
| Visitante | Ninguna | - | - |

CUADRO 6.3: Resultados de apuestas interesante con k-means.

En la figura 6.7 se vio como el grupo 1 está bastante repartido en los 3 posibles resultados, y se conoce que por lo general las cuotas de empate y visitante son mayores que las de los locales, por lo que se podía intuir que establecer una estrategia para el grupo 1 de apostar al local no sería óptimo. En la tabla 6.2 queda corroborado, y es apostar al visitante lo que hubiera dado beneficios. Por lo tanto, según los resultados mostrados en la tabla 6.3, las estrategias que habrá que testear con otra temporada son los mostrados en la tabla 6.4.

| Grupo | Definición de grupo | Apuesta a realizar |
|--------------|----------------------------|---------------------------|
| 1 | Parejo-visitante | Visitante |
| 2 | Parejo-local | Local |
| 3 | Local | Empate |
| 4 | Visitante | Ninguna |

CUADRO 6.4: Apuestas aconsejadas, extraídas de k-means

Como se ha comentado en la teoría, a partir de PC1 PC2 PC3 el método k-means establece unos grupos iniciales, siendo el usuario quien escoge el número de grupos a crear. Cada grupo tendrá un centroide, e iterativamente los grupos irán moviéndose, ya que las observaciones pasarán a formar parte de aquel grupo con distancia euclidiana más cercana al centroide al mismo. El proceso seguirá iterando hasta el punto de converger. En la figura 6.12 se muestran las posiciones de los 4 centroides en la base con ejes principales, como también las distancias entre centroides. Cuando se quiera clasificar una nueva observación, únicamente habrá que calcular a qué centroide se encuentra más cercano. De todas formas esta evaluación se hará mediante el programario Minitab. A partir de las probabilidades y con APC se obtienen las nuevas informaciones en el espacio PC1-PC2-PC3. En esta dimensión es donde se hacen las iteraciones de nuevas muestras y se determinan los centroides.

K-means Cluster Analysis: PC1; PC2; PC3

Final Partition

Number of clusters: 4

| | Number of observations | Within cluster sum of squares | Average distance from centroid | Maximum distance from centroid |
|----------|---------------------------|-------------------------------------|---|---|
| Cluster1 | 160 | 349,548 | 1,287 | 3,135 |
| Cluster2 | 112 | 224,626 | 1,247 | 3,041 |
| Cluster3 | 52 | 157,074 | 1,510 | 3,203 |
| Cluster4 | 52 | 336,383 | 2,241 | 5,866 |

Cluster Centroids

| Variable | Cluster1 | Cluster2 | Cluster3 | Cluster4 | Grand centroid |
|----------|----------|----------|----------|----------|-------------------|
| PC1 | 1,8593 | -1,9125 | -7,0655 | 5,4637 | -0,0000 |
| PC2 | 1,4368 | 1,0505 | -2,6248 | -4,0588 | 0,0000 |
| PC3 | -0,0076 | -0,0024 | 0,0033 | 0,0251 | 0,0000 |

Distances Between Cluster Centroids

| | Cluster1 | Cluster2 | Cluster3 | Cluster4 |
|----------|----------|----------|----------|----------|
| Cluster1 | 0,0000 | 3,7916 | 9,8056 | 6,5723 |
| Cluster2 | 3,7916 | 0,0000 | 6,3294 | 8,9730 |
| Cluster3 | 9,8056 | 6,3294 | 0,0000 | 12,6111 |
| Cluster4 | 6,5723 | 8,9730 | 12,6111 | 0,0000 |

FIGURA 6.12: Salida de Minitab con información de los grupos creados. Observaciones, centroides y distancias entre grupos.

Resultados

Se procedió a realizar los mismos pasos con datos de la temporada siguiente: Pasar de cuotas a probabilidades, reducir dimensionalidad mediante componentes principales, repartir las observaciones en cuatro grupos mediante la técnica clusters k-means, y se aplicó la estrategia definida en el punto anterior a cada uno de los grupos. Los resultados son los mostrados en la tabla 6.5.

| Definición de grupo | Apuesta a realizar | Beneficio | Esperanza matemática |
|---------------------|--------------------|-----------|----------------------|
| Parejo-visitante | Visitante | 33.03 | 1.184 |
| Parejo-local | Local | 2.71 | 1.024 |
| Local | Empate | 18.14 | 1.431 |
| Visitante | Ninguna | - | - |

CUADRO 6.5: Resultados de la fase de pruebas con k-means

Adicionalmente se ha querido aplicar a cada grupo la aplicación de cada estrategia básica (H, D, A), por tal de valorar si fueron las estrategias establecidas las en la temporada de prueba han obtenido mejores resultados o no. En los gráficos 6.12, 6.13, 6.14 y 6.15 se muestran las evoluciones. Para el grupo 1, sí que apostar al visitante ha sido la mejor estrategia. En el grupo 2 apostar al local ha salido rentable y prácticamente en ningún momento supuso pérdidas, todo y que apostar al empate obtuvo mejores resultados, aunque con una tendencia poco estable. Comparando el gráfico 6.9 con 6.14 se observa como la apuesta al local mantiene una cierta estabilidad, hecho que junto a que los resultados son favorables, hace creer que con esta estrategia grandes pérdidas no habrá. En el grupo 3 también es la apuesta al empate la que mejores resultados da, y de la misma forma que en el gráfico 6.10, es una estrategia que sobrevive gracias al éxito continuo y bien cotizado de los empates. Todo y que no se aplicó la estrategia 4, comentar que estrategias H y A salieron rentables, mientras que en la temporada de la que se extrajo el modelo sólo lo fue H. Las tendencias observadas en estos gráficos no permite confiar en la una predicción segura.

1H, 1D y 1A

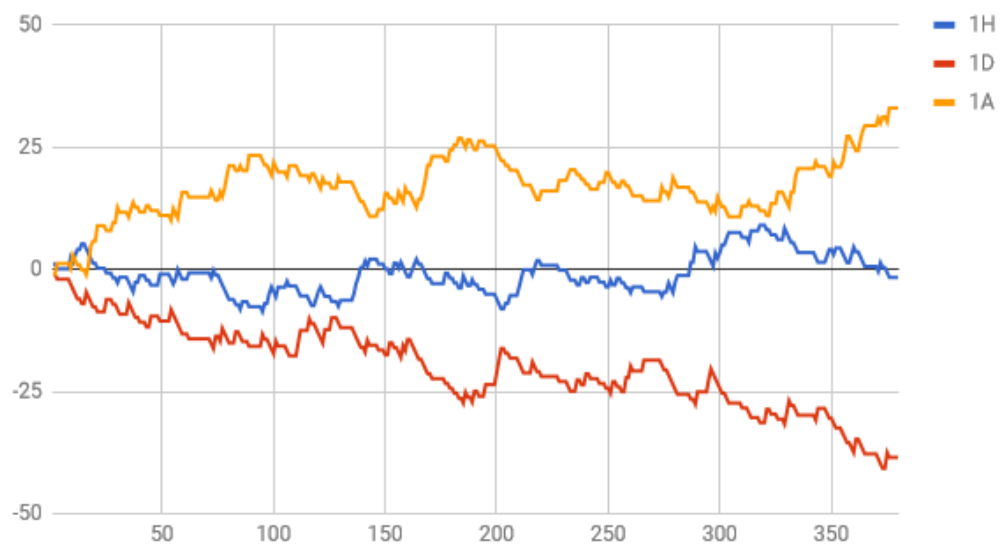


FIGURA 6.13: Evolución de beneficios acumulados apostando únicamente al grupo 1. Fase de pruebas.

2H, 2D y 2A

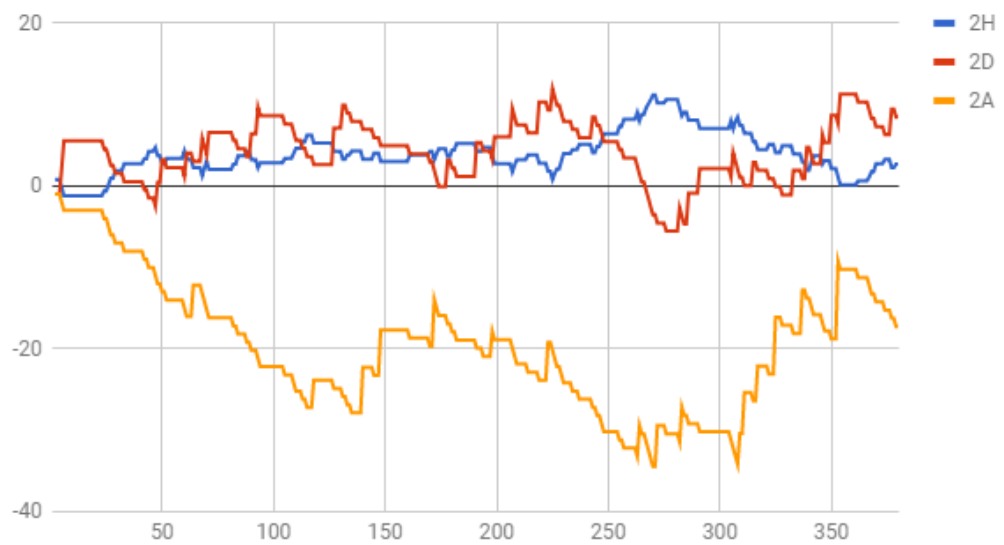


FIGURA 6.14: Evolución de beneficios acumulados apostando únicamente al grupo 2. Fase de pruebas.

3H, 3D y 3A

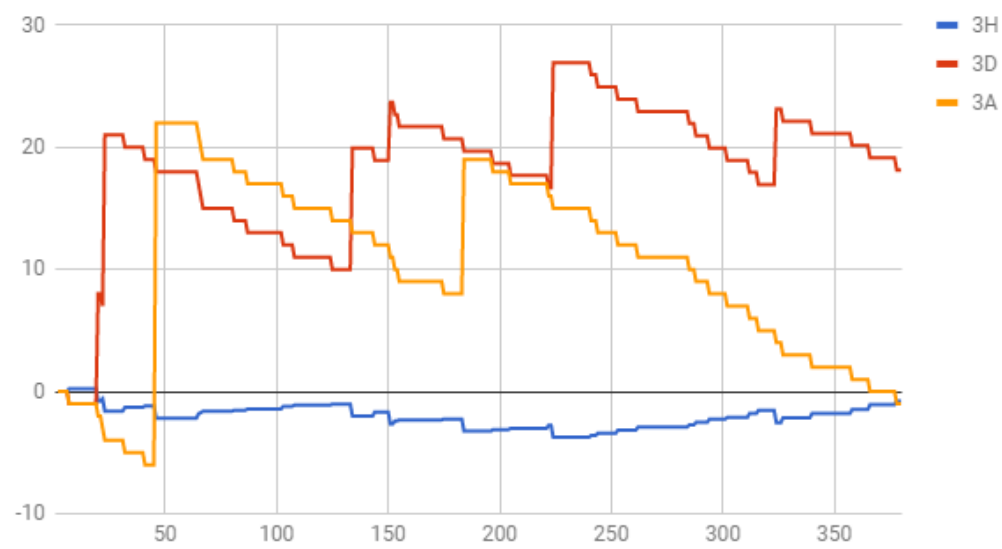


FIGURA 6.15: Evolución de beneficios acumulados apostando únicamente al grupo 3. Fase de pruebas.

4H, 4D y 4A

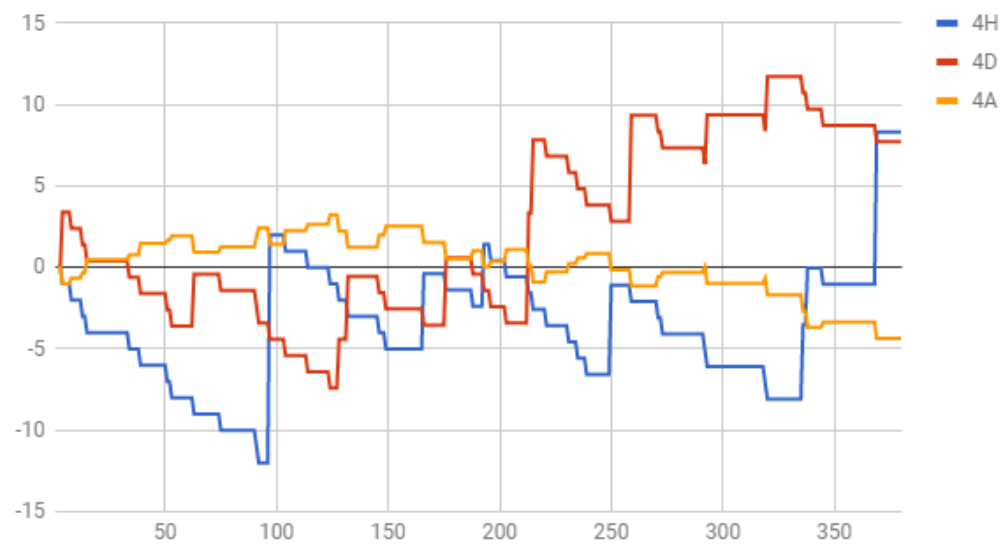


FIGURA 6.16: Evolución de beneficios acumulados apostando únicamente al grupo 4. Fase de pruebas.

6.3. Análisis discriminante

Se hace uso del análisis discriminante para clasificar las observaciones en dos o más grupos cuando ya se tienen muestras con grupos conocidos. También se puede utilizar para investigar de qué manera las variables contribuyen a la separación de grupos.

El proceso de un análisis discriminante puede resumirse en 6 pasos [7]:

- Disponer de un conjunto de datos de entrenamiento (training data) en el que se conoce a qué grupo pertenece cada observación.
- Calcular las probabilidades previas (prior probabilities): la proporción esperada de observaciones que pertenecen a cada grupo.
- Determinar si la varianza o matriz de covarianzas es homogénea en todos los grupos. De esto dependerá que se emplee *LDA* o *QDA*.
- Estimar los parámetros necesarios para las funciones de probabilidad condicional, verificando que se cumplen las condiciones para hacerlo.
- Calcular el resultado de la función discriminante. El resultado de esta determina a qué grupo se asigna cada observación.
- Utilizar validación cruzada (cross-validation) para estimar las probabilidades de clasificaciones erróneas.

Se dispone de análisis discriminante lineal y cuadrático. Con el análisis discriminante lineal, se presupone que todos los grupos poseen la misma matriz de covarianzas. La discriminación cuadrática no parte de este supuesto.

- **Análisis discriminante lineal**

Una observación se clasifica en un grupo si la distancia cuadrada (también llamada distancia de Mahalanobis [7]) de observación hasta el centro del grupo (media) es el mínimo. Se parte del supuesto de que las matrices de covarianzas son iguales para todos los grupos. Existe una parte única de la fórmula de la distancia cuadrada para cada grupo y se denomina función discriminante lineal para ese grupo. Para cualquier observación, el grupo con la distancia cuadrada más pequeña tiene la función discriminante lineal más grande y la observación se clasifica entonces en este grupo.

El análisis discriminante lineal tiene la propiedad de la distancia cuadrada simétrica: la función discriminante lineal del grupo i evaluado con la media del grupo j es igual a la función discriminante lineal del grupo j evaluado con la media del grupo i .

- **Análisis discriminante cuadrático.**

No existen suposiciones con el análisis discriminante cuadrático que indiquen que los grupos tienen matrices de covarianzas iguales. Como con el análisis discriminante

lineal, una observación se clasifica en el grupo que tiene la distancia cuadrada más pequeña. Sin embargo, la distancia cuadrada no se simplifica en una función lineal, de ahí el nombre de análisis discriminante cuadrático.

A diferencia de la distancia lineal, la distancia cuadrática no es simétrica. En otras palabras, la función discriminante cuadrática del grupo i evaluado con la media del grupo j no es igual a la función discriminante cuadrática del grupo j evaluado con la media del grupo i . En consecuencia, la distancia cuadrática es denominada distancia cuadrada generalizada. Si el discriminante de la matriz de covarianza del grupo de la muestra es menor que uno, la distancia cuadrada generalizada puede ser negativa.

El clasificador cuadrático o *Quadratic Discriminant Analysis QDA* se asemeja en gran medida al *LDA*, con la única diferencia de que el *QDA* considera que cada clase k tiene su propia matriz de covarianza.

Paralelamente, se aplica el concepto de validación cruzada:

- Dicha técnica se utiliza para compensar un nivel de significancia aparente optimista. El nivel de significancia aparente es el porcentaje de las observaciones clasificadas erróneamente. Este número tiende a ser optimista porque los datos que se clasifican son iguales a los datos utilizados para construir la función de clasificación [8].

La rutina de la validación cruzada funciona al omitir cada observación una por una, calculando nuevamente la función de clasificación que utilizan los datos restantes y luego clasificando la observación omitida. El tiempo del cálculo es aproximadamente cuatro veces más prolongado con este procedimiento.

6.3.1. Resultados

De nuevo, para el uso de esta técnica se utilizan las componentes principales derivadas de las variables probabilidad, y también la variable respuesta, que es el resultado del partido.

Dado que la técnica consiste en realizar clasificaciones predictivas en base a unas variables predictoras y una variable respuesta, si para la predicción de un caso se utilizase se tuviera en cuenta el propio caso se estaría trabajando con una ventaja que a la práctica no se dispone. Es por ello que se operará con validación cruzada, una variante que permite no tener en cuenta una observación para la predicción de la misma.

Separación lineal

En las figuras 6.17 y 6.18 se puede observar que al utilizar validación cruzada, se aciertan 2 empates menos. Según las tablas de ambas figuras parece que no hay más diferencias que esa, interpretándose esto como que la predicción de una observación va en línea con la

predicción del conjunto de observaciones. Aún así, en la tabla de resultados se verá como sí que existen diferencias en los partidos de visitante, por ejemplo, por lo que los 52 encuentros acertados con validación cruzada y sin no son los mismos en su totalidad.

Filas: Lineal Columnas: FTR

| | A | D | H | Todo |
|------|-----|----|-----|------|
| A | 52 | 17 | 16 | 85 |
| D | 49 | 47 | 68 | 164 |
| H | 8 | 25 | 94 | 127 |
| Todo | 109 | 89 | 178 | 376 |

FIGURA 6.17: Captura de la tabla de correspondencia de los grupos creados y de los grupos reales. Separación lineal

Filas: Lineal Cruzada Columnas: FTR

| | A | D | H | Todo |
|------|-----|----|-----|------|
| A | 52 | 18 | 16 | 86 |
| D | 49 | 45 | 68 | 162 |
| H | 8 | 26 | 94 | 128 |
| Todo | 109 | 89 | 178 | 376 |

FIGURA 6.18: Captura de la tabla de correspondencia de los grupos creados y de los grupos reales. Separación lineal. Validación cruzada.

Aplicando a cada grupo la estrategias básica que le pertoca se llega a los gráficos mostrados en 6.19. Parece que la apuesta a local es la que resulta más rentable, aunque haber supuesto pérdidas durante la primera mitad de temporada hace desconfiar ligeramente en ella. La estrategia de apostar al visitante parece relativamente estable, todo y que con un valor medio que no alcanza valores positivos, no suponiendo beneficios. La tendencia del acumulado de la apuesta al empate no es para nada estable, no transmite seguridad para futuras aplicaciones, además de acabar en considerables pérdidas a final de temporada. En la tabla 6.6 se muestran únicamente los valores finales de beneficio.

H, D y A (Lineal cruzada)

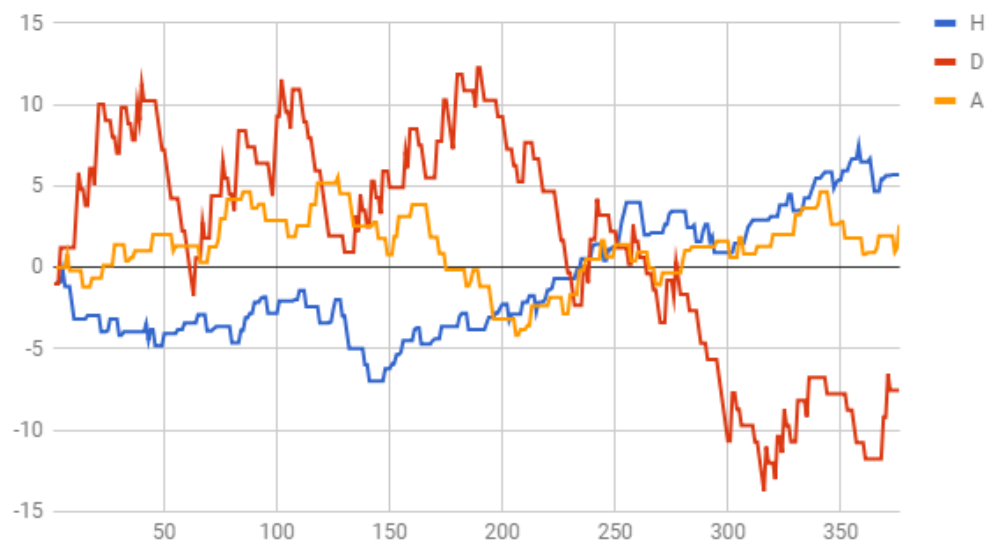


FIGURA 6.19: Evolución de los beneficios acumulados con validación cruzada y separación lineal

| Beneficio | Apostar a local | Apostar a empate | Apostar a visitante |
|------------------------|--------------------|---------------------|------------------------|
| Sin validación cruzada | 6.7 | -2.24 | 3.6 |
| Con validación cruzada | 5.7 | -7.55 | 2.6 |

CUADRO 6.6: Beneficio de las apuestas a los 3 resultados con técnica de análisis discriminante y separación lineal

La apuesta al local es la que interesa, y la apuesta al visitante estaría por ver.

Separación cuadrática

En las figuras 6.20 y 6.21 se muestran las tablas descriptivas, relacionando grupos creados con grupos reales. De nuevo se ve como hay más aciertos sin validación cruzada que con ella.

Filas: Cuadrática Columnas: FTR

| | A | D | H | Todo |
|------|-----|----|-----|------|
| A | 48 | 16 | 15 | 79 |
| D | 50 | 51 | 66 | 167 |
| H | 11 | 22 | 97 | 130 |
| Todo | 109 | 89 | 178 | 376 |

FIGURA 6.20: Captura de la tabla de correspondencia de los grupos creados y de los grupos reales. Separación cuadrática.

Filas: Cuadrática cruzada Columnas: FTR

| | A | D | H | Todo |
|------|-----|----|-----|------|
| A | 47 | 19 | 15 | 81 |
| D | 51 | 48 | 67 | 166 |
| H | 11 | 22 | 96 | 129 |
| Todo | 109 | 89 | 178 | 376 |

FIGURA 6.21: Captura de la tabla de correspondencia de los grupos creados y de los grupos reales. Separación cuadrática. Validación cruzada.

En la figura 6.22 se muestra el acumulado de los beneficios. Apostar a H tiene buena tendencia y será la estrategia escogida. Apostar a A sigue siendo estable, igual que en el modelo con separación lineal, pero acaba alcanzando valores negativos. Apostar a D deja de ser tan inestable, pero no será escogida.

H, D y A (Cuadrática cruzada)

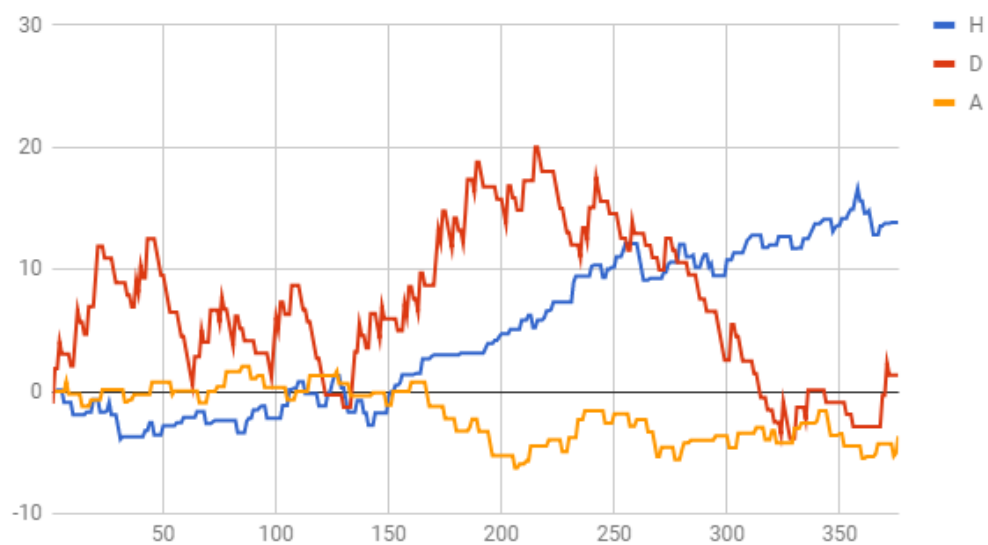


FIGURA 6.22: Evolución de los beneficios acumulados con validación cruzada y separación cuadrática

De nuevo, aplicando las estrategias básicas a cada grupo, los beneficios obtenidos son los mostrados en la tabla 6.7.

| Beneficio | Apostar a local | Apostar a empate | Apostar a visitante |
|------------------------|----------------------------|-----------------------------|--------------------------------|
| Sin validación cruzada | 14.67 | 10.59 | 1.06 |
| Con validación cruzada | 13.84 | 1.34 | -3.64 |

CUADRO 6.7: Beneficio de las apuestas a los 3 resultados con técnica de análisis discriminante y separación cuadrática

| Apuesta a realizar | Beneficio | Participaciones | Esperanza matemática |
|-------------------------------|------------------|------------------------|---------------------------------|
| Local (lineal) | 5.7 | 128 | 1.045 |
| Local (cuadrática) | 13.84 | 129 | 1.107 |

CUADRO 6.8: Beneficio de las apuestas al local con técnica de análisis discriminante y separación cuadrática

- Los resultados sirven de ejemplo para mostrar la importancia de trabajar con validación cruzada. Teniendo en cuenta la propia observación, se clasifican más empates correctamente que sin tenerla en cuenta. De nuevo, parece que la apuesta al local es la que interesa.

Todo y que la validación cruzada ya es probar con nuevos datos, se procede también a probar la estrategia de apuesta a local sobre los partidos de la temporada siguiente. En la figura 6.23 se muestran los acumulados de los beneficios. Parece ser que ambas tendencias son similares, pero con unos resultados más acertado mediante separación cuadrática. No es de extrañar que si los grupos quedan forzados a ser separados por un plano, como sucede en la separación lineal, los resultados sean peores que cuando la separación se realiza con un plano curvado.

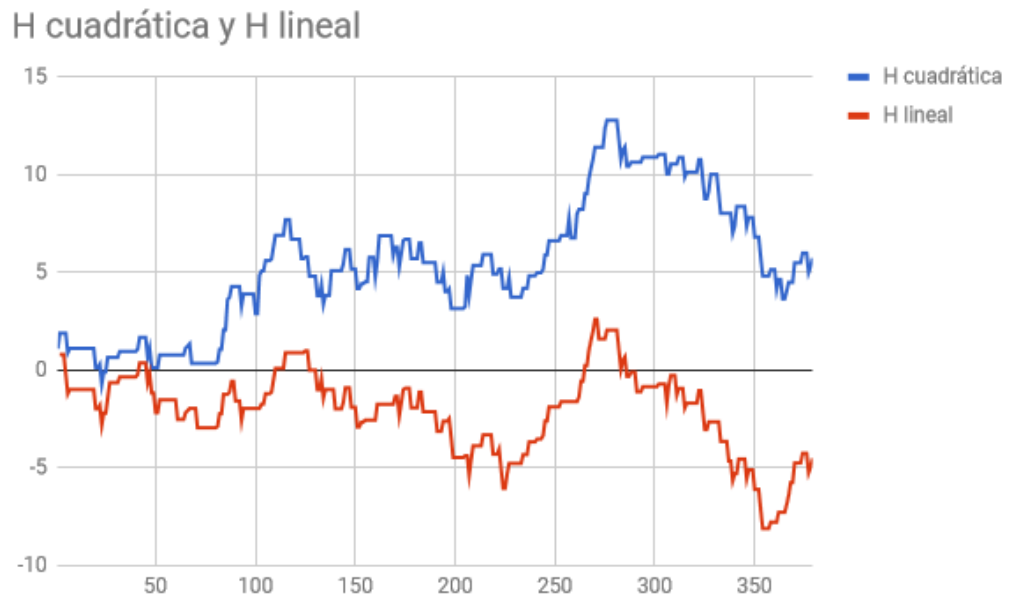


FIGURA 6.23: Evolución de los beneficios acumulados apostando al local con el modelo extraído del análisis discriminante. Fase de pruebas.

Aunque la estrategia aconsejada fue apostar al local, se ha querido evaluar cómo hubieran resultado las estrategias de apostar también con los grupos de empate y visitante. En las figuras 6.24 y 6.25 se muestran dichas evoluciones. En ambas se vuelve a ver que los grupos creados mediante separación lineal dan peores resultados. En la figura 6.25 se ve como la estrategia no empleada de apostar al equipo visitante hubiera resultado rentable. Acudiendo a ver cuáles fueron los precedentes de la apuesta al visitante en la temporada de la que se extrajo el modelo, que se muestran en las figuras 6.19 y 6.22, se ve como realmente no sufrían grandes descensos, pero al no haber alcanzado resultados suficientemente satisfactorios no se fijó como estrategia a seguir. Siendo algo ventajistas, de cara a otras futuras pruebas sí que se tendría en cuenta, por las razones comentadas.

D cuadrática y D lineal

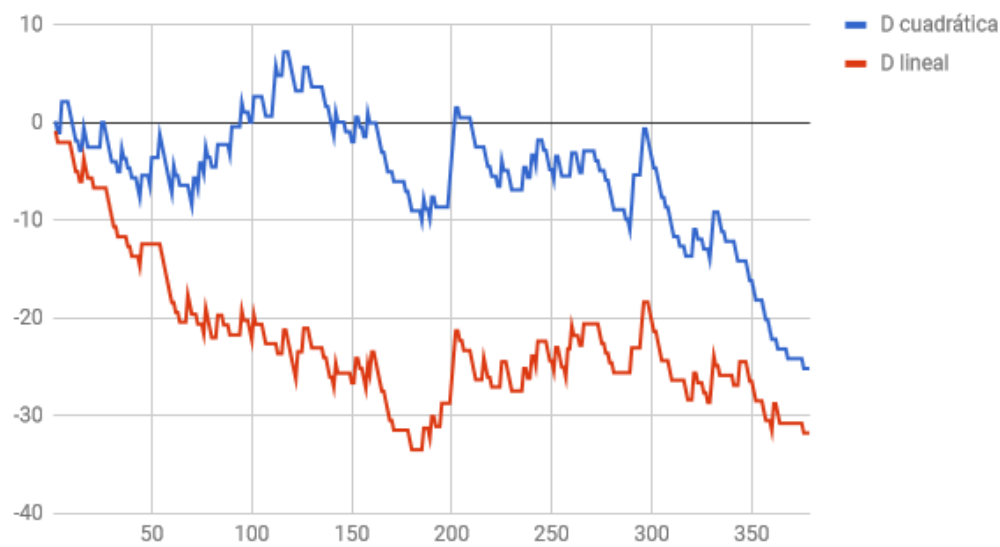


FIGURA 6.24: Evolución de los beneficios acumulados apostando al empate con el modelo extraído del análisis discriminante. Fase de pruebas.

A cuadrática y A lineal

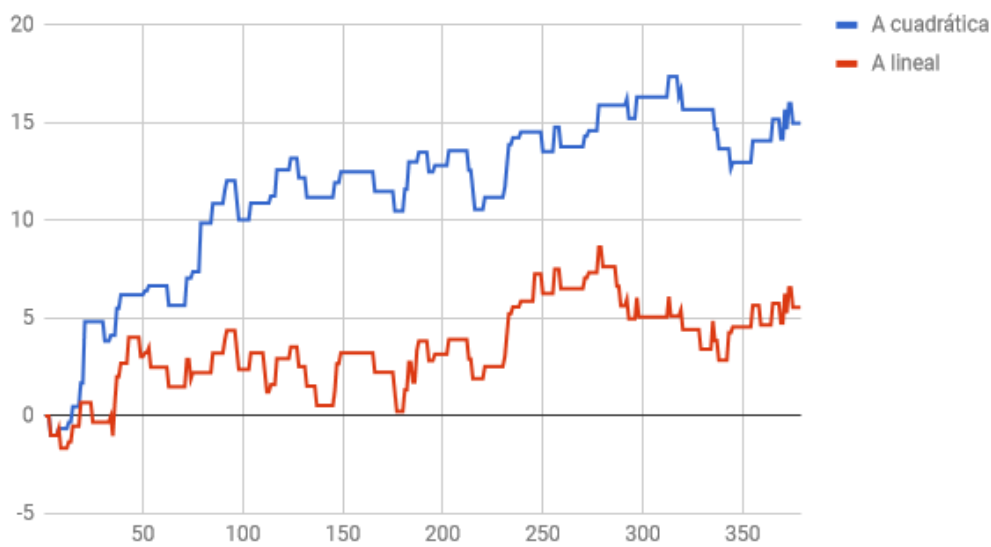


FIGURA 6.25: Evolución de los beneficios acumulados apostando al visitante con el modelo extraído del análisis discriminante. Fase de pruebas.

Conclusiones

Por último, en la tabla 6.9 quedan resumidos los resultados de las estrategias que se fijaron para emplear. Vistos los buenos resultados de la apuestas al visitante y con los precedentes

de que en la temporada de la que se extrajo el modelo se obtuvieron esperanzas matemáticas 10 puntos por encima de la estrategia básicas a visitante, además de una buena tendencia, se ha decidido mostrar también sus resultados también en la tabla 6.9. Los modelos de agrupación cuadrática han demostrado ser más flexibles y tener un ajuste más natural. La importancia de la validación cruzada es la misma que la de testear con datos diferentes, se ha de hacer.

| Apuesta a realizar | Beneficio | Participaciones | Esperanza matemática |
|------------------------|-----------|-----------------|-------------------------|
| Local (lineal) | 4.5 | 131 | 0.965 |
| Local (cuadrática) | 5.75 | 137 | 1.042 |
| Visitante (lineal) | 5.56 | 89 | 1.062 |
| Visitante (cuadrática) | 14.98 | 70 | 1.214 |

CUADRO 6.9: Resumen de las apuestas realizadas y aconsejadas con uso de técnicas de análisis discriminante. Fase de pruebas.

6.4. Regresión Logística Nominal

En la estadística, la regresión lineal es un enfoque lineal para modelar la relación entre una respuesta escalar (o variable dependiente) y una o más variables explicativas (o variables independientes). El caso de una variable explicativa se llama regresión lineal simple. Para más de una variable explicativa, el proceso se llama regresión lineal múltiple. Este término es distinto de la regresión lineal multivariada, donde se predicen variables dependientes correlacionadas múltiples, en lugar de una sola variable escalar.

Respuesta categórica:

Cuando la respuesta es categórica, se hace uso de un tipo de regresión logística denominada multinomial, y que para el caso en que la respuesta categórica tiene dos niveles, también es conocida como regresión logística binomial.

Los logits de las probabilidades binomiales quedan expresados como función lineal de unas variables conocidas x y unos parámetros β . En la ecuación 6.4 se ve su expresión, donde se muestra que está comparando la probabilidad de que el resultado i , π_i venza al resto, que en este caso se trata de sólo una única alternativa.

$$\text{logit}(\pi_i) = \ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \beta_0 + \beta_1 \cdot x_{1,2} + \dots + \beta_k \cdot x_{k,i} \quad (6.4)$$

La presentacia del logaritmo neperiano se debe a que de esta manera cada probabilidad de éxito se encuentra acotada entre los valores 0 y 1, como así debe ser. Cuando el logit alcanza

valores positivos elevados, la probabilidad tiende a 1, y cuando son valores muy negativos tiende a 0. En la ecuación 6.5 se muestra la expresión de la probabilidad de éxito de i .

$$\pi_i = \frac{1}{1 + \exp^{-(\beta_0 + \beta_1 \cdot x_{1,2} + \dots + \beta_k \cdot x_{k,i})}} \quad (6.5)$$

Para el caso multinomial, por parejas de posibles resultados se procede de la misma manera que en la ecuación 6.4. En la ecuación 6.9 se muestran las tres probabilidades asociadas al caso práctico del proyecto. Donde la primer probabilidad está asociada a la victoria local, la segunda al empate y la tercera al visitante. Sin embargo, también se puede ver la comparativa de que el sea más probable un empate que victoria local, en la ecuación 6.6, y la de que la victoria visitante sea más probable que la de local, en la ecuación 6.7.

Se utiliza la regresión logística nominal para realizar una regresión logística en una variable de respuesta nominal utilizando un algoritmo iterativo-reponderado de cuadrados mínimos para obtener los estimados de verosimilitud máxima de los parámetros. Las variables nominales son variables categóricas que tienen tres o más niveles posibles sin ordenamiento natural. Por ejemplo, los niveles en un estudio de degustación pueden incluir crujiente, blando y crocante.

La función de verosimilitud indica la probabilidad de que una muestra observada sea una función de valores de parámetros posibles. Por lo tanto, cuando se maximiza la función de verosimilitud se determina los parámetros que tienen mayor probabilidad de producir los datos observados [8].

Probabilidades:

En la regresión logística nominal y ordinal, una variable de respuesta puede tener tres o más categorías. La probabilidad del evento es la probabilidad de que un patrón dado de factores o covariantes tenga una categoría de respuestas específicas. La probabilidad acumulada de evento es la probabilidad de que la respuesta para un patrón dado de factores o covariantes esté en la categoría k o inferior, para cada k posible, donde k equivale a las categorías de respuesta, $1, \dots, k$. Lo opuesto de un evento es un evento fallido o fracaso.

Para determinar si una variable debe incluirse dentro del modelo de regresión hay que recordar el significado del nivel de significancia, y de las pruebas de hipótesis.

El *p-valor* determina si es adecuado rechazar la hipótesis nula en una prueba de hipótesis. Los valores p abarcan de 0 a 1. El valor p es la probabilidad de obtener una estadística de prueba que sea por lo menos tan extrema como el valor calculado, si la hipótesis nula es verdadera. Antes de realizar cualquier análisis, determine su nivel de significancia (α). Un valor comúnmente utilizado es 0.05. Si el valor p de una estadística de prueba es menor que su nivel de significancia, rechace la hipótesis nula.

Debido a lo indispensable de su función en las pruebas de hipótesis, los valores p se utilizan en muchas áreas de la estadística, entre las que figuran la estadística básica, los modelos lineales, la confiabilidad y los análisis multivariados, entre muchos otros. La clave es comprender lo que representan las hipótesis nulas y alternativas en cada prueba y luego utilizar el valor p para orientarse en su decisión de rechazar la nulidad.

El valor p se calcula a partir de la muestra observada y representa la probabilidad de rechazar incorrectamente la hipótesis nula, cuando en realidad es verdadera (error de tipo I). En otras palabras, es la probabilidad de obtener una diferencia que sea, como mínimo, tan grande como la diferencia entre el valor observado y el valor hipotético a través del error aleatorio solamente.

6.4.1. Resultados

Se comienza con la búsqueda de un modelo de regresión, incorporando todas las variables, pero como se muestra en la figura 6.26, no todas las variables son significativas. Cuando el valor de p es mayor que 0.05 no se puede asegurar que esa variable tenga un coeficiente diferente de cero. Es por ello que se eliminarán aquellas no significativas, una por una y probando el nuevo modelo, hasta llegar a aquél que cumpla la condición, y que es el mostrado en la figura 6.27.

Tabla de regresión logística

| Predictor | Coef | SE Coef | Z | P | Relación de probabilidades | IC de 95% Inferior | Superior |
|----------------|------------|-----------|-------|-------|----------------------------|--------------------|----------|
| Logit 1: (D/H) | | | | | | | |
| Constante | -0,591569 | 0,143838 | -4,11 | 0,000 | | | |
| PC1 | 0,183339 | 0,0421843 | 4,35 | 0,000 | 1,20 | 1,11 | 1,30 |
| PC2 | 0,0298937 | 0,0732748 | 0,41 | 0,683 | 1,03 | 0,89 | 1,19 |
| PC3 | -0,129766 | 0,726126 | -0,18 | 0,858 | 0,88 | 0,21 | 3,65 |
| Logit 2: (A/H) | | | | | | | |
| Constante | -0,716782 | 0,172101 | -4,16 | 0,000 | | | |
| PC1 | 0,400709 | 0,0550237 | 7,28 | 0,000 | 1,49 | 1,34 | 1,66 |
| PC2 | -0,0661170 | 0,0735112 | -0,90 | 0,368 | 0,94 | 0,81 | 1,08 |
| PC3 | -0,566768 | 0,797285 | -0,71 | 0,477 | 0,57 | 0,12 | 2,71 |

FIGURA 6.26: Tabla de regresión logística con todas las variables

Tabla de regresión logística

| Predictor | Coef | SE Coef | Z | P | Relación de probabilidades | IC de 95% Inferior | Superior |
|----------------|-----------|-----------|-------|-------|----------------------------|--------------------|----------|
| Logit 1: (D/H) | | | | | | | |
| Constante | -0,575111 | 0,135926 | -4,23 | 0,000 | | | |
| PC1 | 0,180562 | 0,0399613 | 4,52 | 0,000 | 1,20 | 1,11 | 1,30 |
| Logit 2: (A/H) | | | | | | | |
| Constante | -0,786971 | 0,162216 | -4,85 | 0,000 | | | |
| PC1 | 0,434542 | 0,0519010 | 8,37 | 0,000 | 1,54 | 1,39 | 1,71 |

FIGURA 6.27: Tabla de regresión logística con las variables significativas

Encontrado el modelo, se hará uso de una opción con la que se obtendrán 3 columnas, correspondientes a la probabilidad de éxito de cada resultado según este modelo de regresión.

La estrategia que se va a aplicar será aplicable a cada posible resultado. Se comparará la probabilidad obtenida con el modelo, a la probabilidad más alta extraída a partir de las cuotas, y si en algún caso la del modelo es superior se realizará la apuesta. En otras palabras, si el modelo predice que un resultado es más probable que lo que marcan las cuotas, es momento de apostar.

En la gráfica 6.28 se muestran las evoluciones de los beneficios acumulados aplicando cada una de las estrategias y bajo las condiciones estipuladas. En la tabla 6.10 se muestran los resultados obtenidos, y el número de ocasiones en que la probabilidad encontrada por el modelo fue superior a la marcada por las cuotas:

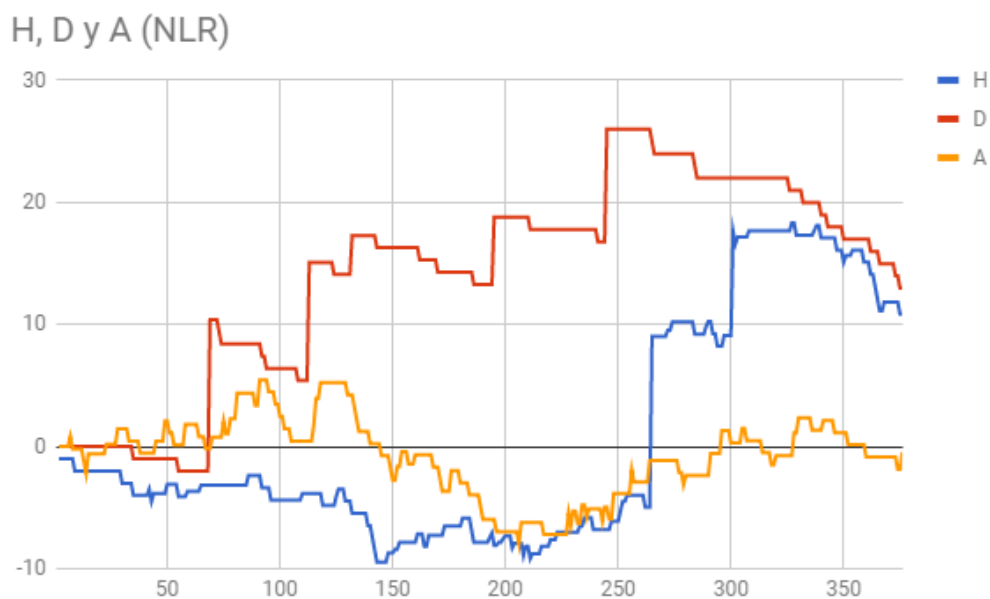


FIGURA 6.28: Evolución de los beneficios acumulados aplicando la estrategia obtenida con NLR

| Apuesta a realizar | Beneficio | Participaciones | Esperanza matemática |
|-----------------------|-----------|-----------------|-------------------------|
| Local | 10.84 | 86 | 1.126 |
| Empate | 12.98 | 32 | 1.406 |
| Visitante | -0.46 | 89 | 0.994 |

CUADRO 6.10: Resumen de la estrategia implementada con NLR

Dado que ninguna ganancia es claramente desfavorable, se aplicarán las 3 estrategias básicas en la siguiente temporada. Para determinar las probabilidades de la temporada de prueba se aplicará la ecuación 6.9, encontrada a partir de la demostración, donde la probabilidad se menciona como Π .

$$\log \left(\frac{\pi_D}{\pi_H} \right) = \beta_0^1 + \beta_1^1 \cdot PC1 = \hat{y}^1 \quad (6.6)$$

$$\log \left(\frac{\pi_A}{\pi_H} \right) = \beta_0^2 + \beta_1^2 \cdot PC1 = \hat{y}^2 \quad (6.7)$$

$$\pi_H + \pi_D + \pi_A = 1 \quad (6.8)$$

$$\left(\frac{1}{1 + \hat{y}^1 + \hat{y}^2}, \hat{\pi}_H \cdot e^{y^{(1)}}, \hat{\pi}_H \cdot e^{y^{(2)}} \right) = (\pi_H, \pi_D, \pi_A) \quad (6.9)$$

En el gráfico 6.29 se muestran las evoluciones en la temporada de prueba de test, y en la tabla 6.11 el resumen de los resultados.

H, D y A (NLR test)

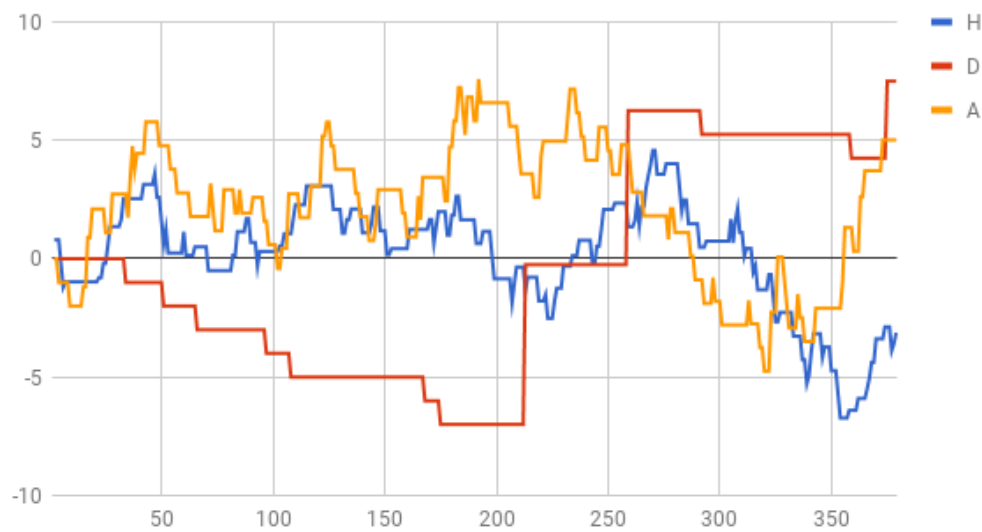


FIGURA 6.29: Evolución de los beneficios acumulados aplicando la estrategia obtenida con NLR. Fase de pruebas.

| Apuesta a realizar | Beneficio | Participaciones | Esperanza matemática |
|--------------------|-----------|-----------------|----------------------|
| Local | -3.13 | 125 | 0.975 |
| Empate | 7.5 | 12 | 1.625 |
| Visitante | 5.02 | 102 | 1.049 |

CUADRO 6.11: Resumen de la estrategia implementada con NLR. Fase de pruebas.

Probabilidad corregida

Como ya se mostró en el apartado 4.3, 4.3, la transformación de cuota a probabilidad lleva a una suma de probabilidades superiores a cien. Es por ello que se propuso una manera de recalcular las probabilidades, de forma que sumaran cien. Se procede a realizar lo mismo con las probabilidades de las cuotas que se utilizan en esta estrategia. La ventaja será que se estará comparando con las probabilidades reales que tiene una de las partes a tener éxito, según dicen las cuotas. En la figura 6.30 se muestra el entrenamiento, y en la tabla 6.12 los resultados.

H, D y A (NLR). Corrección.

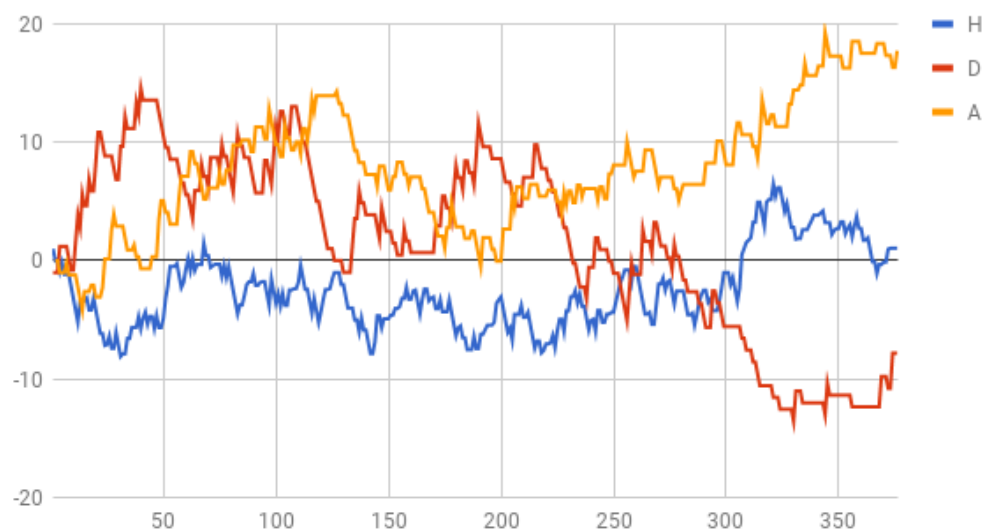


FIGURA 6.30: Evolución de los beneficios acumulados aplicando la estrategia obtenida con NLR

| Apuesta a realizar | Beneficio | Participaciones | Esperanza matemática |
|--------------------|-----------|-----------------|----------------------|
| Local | 1.05 | 222 | 1.005 |
| Empate | -7.8 | 84 | 0.907 |
| Visitante | 17.74 | 157 | 1.113 |

CUADRO 6.12: Resumen de la estrategia implementada con NLR. Corrección.

El empate no parece suficiente estable y además acaba con pérdidas considerables. La victoria local evoluciona en fases negativas, y la victoria visitante en positivas. Es por ello que únicamente ésta última es la que se recomienda para la fase de pruebas.

Pese a sólo recomendar la del visitante, se prueban las 3 estrategias. En la figura se muestra la evolución de todas ellas y en la tabla 6.13 el resumen.

| Apuesta a realizar | Beneficio | Participaciones | Esperanza matemática |
|--------------------|-----------|-----------------|----------------------|
| Visitante | 19.87 | 157 | 1.126 |

CUADRO 6.13: Resumen de la estrategia implementada con NLR con corrección. Fase de pruebas.

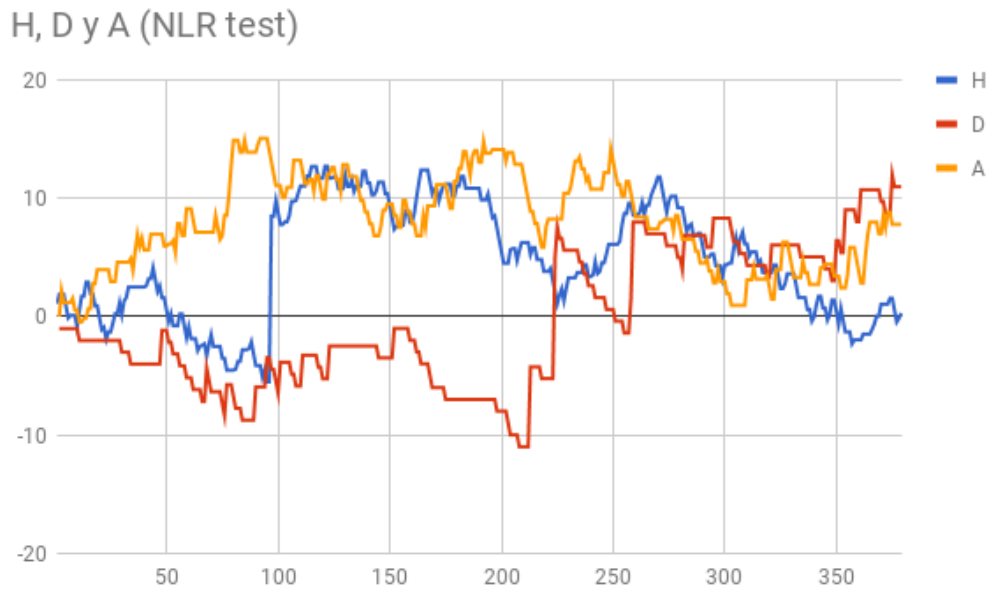


FIGURA 6.31: Evolución de los beneficios acumulados aplicando la estrategia obtenida con NLR. Fase de pruebas.

Se puede decir que el comportamiento de la evolución del visitante en training y test son similares. Realmente la hipotética apuesta al empate hubiera sido rentable, pero por como fue su evolución en el entrenamiento no se fijó.

Conclusiones

A diferencia de los anteriores modelos, los resultados de la temporada con el modelo y de la temporada a aplicar las estrategias, los resultados no son demasiado acordes. Quizá faltó valorar en el gráfico 6.28 si la apuesta a H era interesante, pues ciertamente parece que la estrategia dio un buen resultado final gracias a 2 partidos en que ganó el equipo local contra un equipo puntero. Dependiendo de si quien establece la estrategia acepta estas observaciones como ocasionalmente normales o si no. En la temporada de pruebas se ve como los resultados de esta estrategia no son los deseados. Respecto a la estrategia del empate, en ambas temporadas ha resultado interesante realizar esta apuesta, pero hay que valorar que la participación es baja, y de nuevo los buenos resultados son a base de observaciones puntuales bien cotizadas. El modelo de regresión no ha ofrecido resultados tan satisfactorios como las técnicas que emplean grupos.

Haber añadido la misma estrategia pero con las probabilidades corregidas hace el problema más real. Se pierde la relación probabilidad-cuota, pero la comparación que se realiza es la correcta. La estrategia que se escogió, de apuesta al visitante, mantuvo unos patrones y acabó dando beneficios.

Capítulo 7

Planificación

En la labor de ingeniero no podría faltar mostrar un resumen de la planificación, y se mostrará en forma de diagrama de Gantt. Esta herramienta permite visualizar los tiempos previstos para realizar las diferentes tareas. A cada tarea se le asigna un distintivo, que en este caso será un dígito y una letra, que corresponden a la fase del proyecto y el orden dentro de fases, respectivamente. Se ha añadido dos columnas con las horas dedicadas a cada tarea, diferenciándolas según si serían tratadas por un analista senior o por un ingeniero novel. Dentro del anexo B se muestra el diagrama de Gantt teórico, en las imágenes B.1, B.2 y B.3 .

Dado que el proyecto se ha realizado paralelamente con otras obligaciones personales, en ocasiones los tiempos no se han cumplido, además que las horas fijadas a una actividad según planning en gran parte de las tareas no se ha podido cumplir. En el anexo B también se muestra un diagrama de Gantt de una aproximación de lo realizado. Queda adjunta en las figuras B.4, B.5, B.6 y B.7. Cabe decir que esta herramienta se utiliza al iniciar el proyecto, pero ha parecido buena opción tomarla también para mostrar una aproximación tiempos reales.

Capítulo 8

Presupuesto

De nuevo, forma parte de un proyecto dejar reflejado una idea de qué recursos se necesitaron y en cuánto se pueden evaluar económicamente, y por tanto también el coste del proyecto. En la tabla 8.1 se muestran los recursos de los que se hizo uso y el coste que suponía cada uno de ellos y en la 8.2 el presupuesto del proyecto.

Se fijó en 12€/h el salario de un analista senior documentandose sobre esta información, y en 8€/h el salario de un ingeniero novel porque éste es el valor orientativo que aconseja UPC. Dado que el convenio laboral estipula que se deben realizar 30 minutos de descanso a partir de las 4 horas y media trabajadas, aquí se han tenido en cuenta también. Se han considerado jornadas laborables de 6h, por lo que por cada 5.5h hay 0.5 de descanso. Las tarifas de coste de potencia eléctrica varían oscilan alrededor de 0.12€/kWh y por ello se ha considerado este valor. Al ser estudiante UPC se ha podido disponer de una licencia de Minitab de forma gratuita, y se ha podido adquirir el paquete MicrosoftOffice por un precio más accesible. Por contra, por haber cursado el trabajo de fin de grado, también se ha tenido que desembolsar la cantidad correspondiente a los créditos cursados. El programa TeXstudio también se ha adquirido de manera gratuita.

| Recursos utilizados | Coste |
|------------------------------------|--|
| 300h laborables | 12€/h(analista) - 8€/h(ingeniero becado) |
| 300h portátil a potencia media 60W | 0.12€/kWh |
| 150h flexo a potencia media 10W | 0.12€/kWh |
| 27h descansos | 12€/h(analista) - 8€/h(ingeniero becado) |
| Folios y material de escritura | Valorado en 10€ |
| 5 meses portátil (amortización) | 800€/lustró |
| 5 meses MicrosoftOffice | 80€/año |
| 5 meses Minitab | Gratuito |
| 1 mes TeXstudio | Gratuito |
| Estudiante UPC (12 créditos) | 35.77€/crédito |

CUADRO 8.1: Recursos utilizados y costes asociados

| Recursos | Tiempo [h] | €/h | Total [€] |
|--|------------|-----------|-----------|
| Trabajo | | | |
| • Analista | 58 | 12 | 696 |
| • Ingeniero | 242 | 8 | 1936 |
| • Descansos estipulados | 27 | 8.86 | 239.33 |
| Equipo | | | |
| • Gasto energético portátil | 300 | - | 2.16 |
| • Gasto energético flexo | 150 | - | 0.18 |
| • Folios y material de escritura | - | - | 10 |
| • Amortización portatil | - | - | 66.67 |
| • 5 meses Microsoft Office ® license | - | - | 33.33 |
| • Estudiante UPC (12 créditos) ® license | - | - | 429.29 |
| | | Total [€] | 3413 |

CUADRO 8.2: Resumen de costes estimados

Capítulo 9

Conclusiones

- Se ha conseguido diseñar varias estrategias que superan las expectativas.
- Las técnicas que conocían los grupos finales, análisis discriminante y regresión logística nominal, no tuvieron tan buenos resultados como agrupación por k-Means, quizá porque las anomalías modifican un modelo que en fase de pruebas daría mejores resultados.
- La caracterización de una casa de apuestas resultó satisfactoria. Dada la alta posibilidad de estar cometiendo sobreajuste, se probó con varias temporadas más y los resultados claramente favorables se mantuvieron.
- Aunque no se hubiesen encontrado estrategias que generaran beneficios, se hubiera valorado el trabajo de manera notable dado que se ha conseguido implementar varias herramientas de análisis de datos, procediendo con rigurosidad y apoyando al desarrollo personal.
- Los resultados de las fases de training solieron ser mejores que los de test, como era de esperar, y confirmando así el buen uso de probar las estrategias a otros datos diferentes.
- Las casas de apuestas tienen algoritmos de mayor complejidad que lo que este proyecto pudiera abarcar, por lo que resultaba más razonable aprovechar esta información, que no intentarla vencer, diseñando modelos predictivos que hicieran lo mismo, por ejemplo.
- El no haber dado tanta importancia al porcentaje de aciertos se ha considerado acertado, ya que las estrategias que se plantearon fueron a largo plazo. Aún así, sí que se considera importante una variable que no se ha mostrado explícitamente en todo momento, como es el número de participaciones. Cuando un modelo se obtiene con una participación más elevada se puede confiar más en él, igual que cuando las tendencias del desarrollo de la temporada se consideran estables.

- Por el alcance del proyecto y los límites que se fijaron, únicamente se ha operado con cuotas previas al partido. Aún así, como proyecto de mejora se aconseja hacer un intento por trabajar con las cuotas de los últimos 3 días de un partido, pues estas fluctúan. Se realizaron unas recopilaciones de datos y debido a la no coordinación entre casa de apuestas, es más fácil detectar partidos con surebet, como también se piensa que se podrían realizar series temporales para intentar anticiparse a valores máximos de cuotas.
- Se cree que sí se podrá aconsejar a alguien sin conocimientos en fútbol y/o apuestas de fútbol y obtener beneficios.
- A modo de resumen, aunque algo inusual en unas conclusiones, se ha querido mostrar en la tabla 9.1 un resumen de las esperanzas matemáticas, en la figura 9.1 la evolución de las estrategias planteadas como referencia, y en la figura 9.2 las evoluciones de las estrategias diseñadas.
- Finalmente, sí que se han obtenido unas estrategias que se pueden compartir a alguien con desconocimiento de fútbol y puede obtener beneficios.

| Apuesta | Estrategias básicas | k-Means | | Análisis discriminate | | NLR (A*: corrección) | | Caracterización | |
|---------|---------------------|----------|-------|-----------------------|-------|----------------------|--------|-----------------|------|
| | | Training | Test | Training | Test | Training | Test | Training | Test |
| H | 0.95 | 1.117 | 1.024 | 1.107 | 1.042 | 1.126 | 0.975 | 1.39 | 2.09 |
| D | 0.91 | 1.273 | 1.431 | - | - | 1.406 | 1.625 | - | - |
| A | 0.86 | 1.085 | 1.184 | 0.955 | 1.214 | 1.113* | 1.126* | - | - |

CUADRO 9.1: Resumen de las mejoras conseguidas en forma de esperanza matemática

Situación inicial: Estrategias básicas

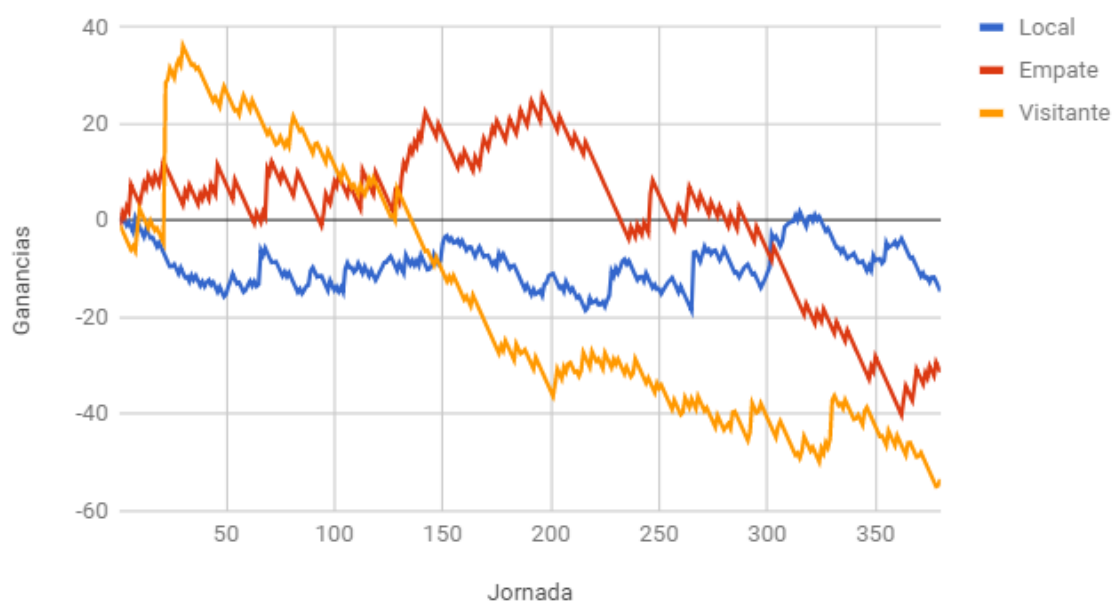


FIGURA 9.1: Evolución de las estrategias básicas

Situación final: Estrategias aconsejadas y probadas

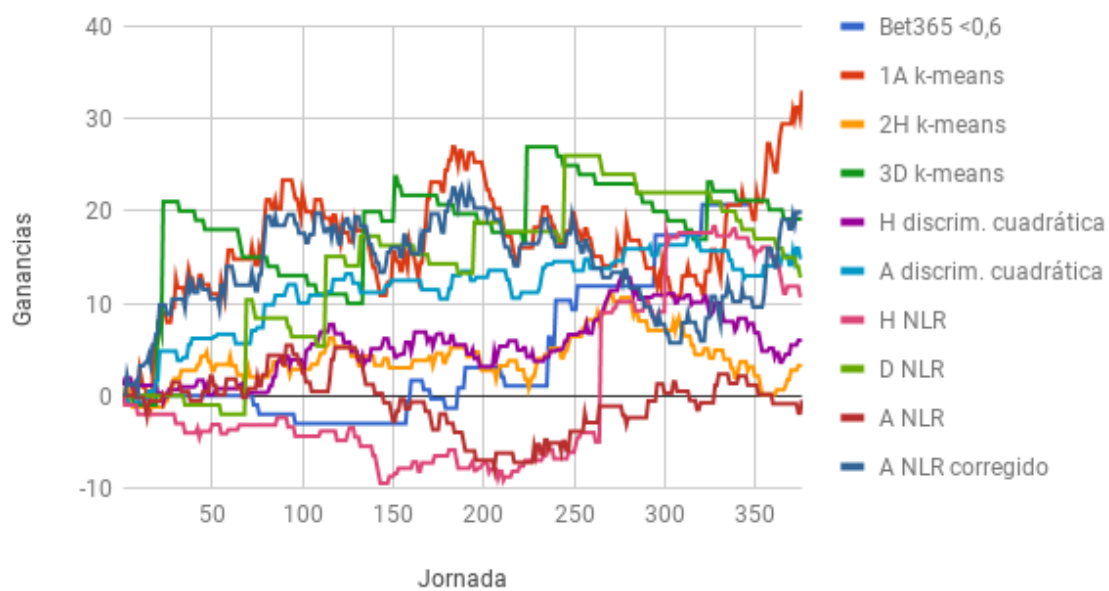


FIGURA 9.2: Evolución de las estrategias diseñadas

Capítulo 10

Impacto medioambiental

Personalmente, es de obligación moral actuar teniendo en cuenta la huella que uno deja. Como proyectista, es de obligación también legal cumplir la legislación medioambiental.

Se han usado alrededor de 70 folios durante el transcurso del proyecto, mayormente nuevas y no recicladas. El proyecto se ha alargado a unas 350h, las cuáles se podrían considerar siempre con uso de ordenador personal conectado a la red eléctrica, además de unas 150-200h de luz artificial. En teoría, el origen de dicha energía no es de combustibles fósiles, ni de otras fuentes no renovables. Sin embargo, hay una pequeña componente de emisiones. Sin embargo, durante toda la suma de horas se ha trabajado conectados a internet mediante Wi-Fi, la cual sí contribuye a la contaminación electromagnética.

En resumen, el tipo de trabajo no ha dado opción a contaminar y en el perímetro de acción se ha actuado correctamente.

Capítulo 11

Agradecimientos

Acabado el trabajo, quiero mostrarme agradecido con todo aquél que de alguna manera me haya aportado en el desarrollo del mismo. Por cuestiones de guía y tutoría, a mis tutores, Lesly María Acosta y Josep Antón Sànchez. Por los conocimientos adquiridos en una optativa vital para un buen desenlace del trabajo, a Josep Ginebra. También a la facultad, gracias a la que he podido crecer. Por apoyo y ánimo, a mi familia, amigos y compañeros más apegados. Por haberme hecho llegar hasta aquí con la motivación de encarar este tema, gracias a mis dos amigos y compañeros con los que hace ya años empezamos a curiosear estos temas. Ha sido un placer haber podido trabajar los últimos meses del grado en este proyecto.

Bibliografía

- [1] J. Buchdahl. (2-2018). Football Data, dirección: <http://www.football-data.co.uk/spainm.php>.
- [2] Minitab18. (2-2018). Methods and formulas for Principal Components Analysis, dirección: <https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/principal-components/methods-and-formulas/methods-and-formulas/>.
- [3] ImagenWikipedia. (6-2018). GaussianScatterPCA, dirección: https://es.wikipedia.org/wiki/An%C3%A1lisis_de_componentes_principales#/media/File:GaussianScatterPCA.svg.
- [4] S. de la Fuente Fernández. (5-2018). Componentes Principales APC, dirección: <http://www.fuenterrebollo.com/Economicas/ECONOMETRIA/MULTIVARIANTE/ACP/ACP.pdf>.
- [5] Aprizze. (5-2018). Microsoft Azure Machine Learning Chapter 8. Clustering, dirección: http://aprizze.info/microsoft/azure_1/8.html.
- [6] V. Lavrenko. (5-2018). K-means clustering: how it works, dirección: https://www.youtube.com/watch?v=_aWzGGNrcic.
- [7] J. A. Rodrigo. (5-2018). Análisis discriminante lineal (LDA) y Análisis discriminante cuadrático (QDA), dirección: https://rpubs.com/Joaquin_AR/233932.
- [8] Minitab18. (5-2018). Distance and discriminant functions for Discriminant Analysis, dirección: <https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/discriminant-analysis/methods-and-formulas/distance-and-discriminant-functions/>.

Apéndice A

Base de datos

| Date | HomeTeam | AwayTeam | FTHG | FTAG | FTR | HTHG | HTAG | HTR | HS | AS | HST | AST | HF | AF | HC | HY | AY | HR | AR | B365H | B365D | B365A | BWH | BWD | BWA | |
|-----------|-------------|-------------|------|------|-----|------|------|-----|----|----|-----|-----|----|----|----|----|----|----|----|-------|-------|-------|------|------|------|------|
| 8/19/2016 | La Coruna | Eibar | 2 | 1 | H | 0 | 0 | 0 | 11 | 12 | 6 | 6 | 14 | 17 | 1 | 8 | 3 | 0 | 0 | 1.95 | 3.4 | 4.2 | 1.95 | 3.3 | 4.2 | |
| 8/19/2016 | Malaga | Osasuna | 1 | 1 | D | 0 | 0 | 0 | 14 | 8 | 4 | 3 | 11 | 22 | 5 | 3 | 2 | 0 | 0 | 1.62 | 3.75 | 6 | 1.55 | 3.5 | 6.25 | |
| 8/20/2016 | Barcelona | Betis | 6 | 2 | H | 3 | 1 | H | 30 | 6 | 12 | 2 | 17 | 13 | 7 | 2 | 1 | 3 | 0 | 1.08 | 11 | 29 | 1.09 | 11 | 23 | |
| 8/20/2016 | Granada | Villarreal | 1 | 1 | D | 0 | 0 | 0 | 7 | 12 | 1 | 3 | 17 | 11 | 7 | 3 | 2 | 0 | 0 | 2.5 | 3.2 | 3 | 2.55 | 3.1 | 3 | |
| 8/20/2016 | Sevilla | Espanol | 6 | 4 | H | 3 | 3 | D | 16 | 10 | 11 | 7 | 8 | 15 | 8 | 2 | 2 | 0 | 0 | 1.67 | 4 | 5 | 1.67 | 3.9 | 5.25 | |
| 8/21/2016 | Ath Madrid | Alaves | 1 | 1 | D | 0 | 0 | 0 | 27 | 2 | 6 | 1 | 7 | 15 | 20 | 2 | 4 | 0 | 0 | 1.2 | 6.5 | 17 | 1.2 | 6 | 18.5 | |
| 8/21/2016 | Sociedad | Real Madrid | 0 | 3 | A | 0 | 2 | A | 8 | 18 | 0 | 8 | 22 | 11 | 4 | 7 | 2 | 3 | 0 | 4.75 | 3.8 | 1.75 | 5 | 3.75 | 1.72 | |
| 8/21/2016 | Sp Gijon | Ath Bilbao | 2 | 1 | H | 0 | 0 | 0 | 12 | 13 | 4 | 4 | 18 | 15 | 2 | 5 | 3 | 0 | 0 | 4 | 3.5 | 1.95 | 4.2 | 3.5 | 1.91 | |
| 8/22/2016 | Celta | Leganes | 0 | 1 | A | 0 | 0 | 0 | 10 | 15 | 2 | 6 | 5 | 13 | 8 | 4 | 1 | 0 | 1 | 1.65 | 3.75 | 5.75 | 1.57 | 3.9 | 6.25 | |
| 8/22/2016 | Valencia | Las Palmas | 2 | 4 | A | 2 | 3 | A | 17 | 6 | 8 | 4 | 15 | 18 | 12 | 1 | 2 | 0 | 0 | 1.73 | 3.75 | 5 | 1.7 | 3.8 | 5 | |
| 8/26/2016 | Betis | La Coruna | 0 | 0 | D | 0 | 0 | 0 | 15 | 12 | 5 | 1 | 20 | 15 | 13 | 2 | 0 | 0 | 0 | 2.2 | 3.2 | 3.5 | 2.2 | 3.3 | 3.4 | |
| 8/26/2016 | Espanol | Malaga | 2 | 2 | D | 1 | 0 | H | 7 | 9 | 2 | 3 | 20 | 17 | 3 | 4 | 2 | 4 | 0 | 2.3 | 3.1 | 3.4 | 2.25 | 3.25 | 3.4 | |
| 8/27/2016 | Eibar | Valencia | 1 | 0 | H | 0 | 0 | 0 | 13 | 13 | 6 | 3 | 13 | 11 | 6 | 10 | 2 | 2 | 0 | 2.7 | 3.2 | 2.75 | 2.6 | 3.2 | 2.8 | |
| 8/27/2016 | Leganes | Ath Madrid | 0 | 0 | D | 0 | 0 | 0 | 6 | 10 | 1 | 3 | 19 | 5 | 1 | 8 | 3 | 1 | 0 | 8.5 | 3.6 | 1.53 | 7 | 3.8 | 1.55 | |
| 8/27/2016 | Osasuna | Sociedad | 0 | 2 | A | 0 | 1 | A | 8 | 5 | 3 | 4 | 14 | 14 | 10 | 1 | 2 | 1 | 0 | 3.1 | 3.1 | 2.5 | 3 | 3.2 | 2.5 | |
| 8/27/2016 | Real Madrid | Celta | 2 | 1 | H | 0 | 0 | 0 | 24 | 8 | 10 | 3 | 8 | 11 | 7 | 2 | 0 | 0 | 0 | 1.18 | 7 | 17 | 1.2 | 7 | 13 | |
| 8/28/2016 | Alaves | Sp Gijon | 0 | 0 | D | 0 | 0 | 0 | 13 | 7 | 2 | 1 | 10 | 12 | 4 | 6 | 1 | 1 | 0 | 2.15 | 3.2 | 3.75 | 2.15 | 3.25 | 3.6 | |
| 8/28/2016 | Ath Bilbao | Barcelona | 0 | 1 | A | 0 | 1 | A | 9 | 12 | 4 | 3 | 20 | 13 | 4 | 4 | 5 | 4 | 0 | 6.5 | 4.33 | 1.5 | 6.5 | 4.33 | 1.5 | |
| 8/28/2016 | Las Palmas | Granada | 5 | 1 | H | 1 | 1 | D | 15 | 13 | 10 | 2 | 8 | 16 | 6 | 4 | 5 | 1 | 0 | 1.91 | 3.4 | 4.33 | 1.9 | 3.5 | 4.2 | |
| 8/28/2016 | Villarreal | Sevilla | 0 | 0 | D | 0 | 0 | 0 | 12 | 6 | 5 | 0 | 16 | 12 | 5 | 1 | 1 | 3 | 0 | 2.5 | 3.4 | 2.8 | 2.55 | 3.4 | 2.75 | |
| 9/9/2016 | Sociedad | Espanol | 1 | 1 | D | 0 | 0 | 0 | 9 | 4 | 3 | 1 | 10 | 10 | 11 | 1 | 2 | 2 | 0 | 1 | 2.05 | 3.4 | 3.75 | 2 | 3.4 | 3.9 |
| 9/10/2016 | Barcelona | Alaves | 1 | 2 | A | 0 | 1 | A | 16 | 5 | 2 | 4 | 12 | 12 | 13 | 0 | 1 | 1 | 0 | 1.07 | 12 | 34 | 1.07 | 11.5 | 36 | |
| 9/10/2016 | Celta | Ath Madrid | 0 | 4 | A | 0 | 0 | 0 | 9 | 13 | 1 | 6 | 13 | 12 | 6 | 4 | 1 | 2 | 0 | 6 | 3.3 | 1.73 | 5.5 | 3.6 | 1.7 | |
| 9/10/2016 | Malaga | Villarreal | 0 | 2 | A | 0 | 2 | A | 15 | 12 | 2 | 6 | 8 | 14 | 9 | 7 | 2 | 1 | 0 | 2.3 | 3.1 | 3.3 | 2.4 | 3.1 | 3.25 | |
| 9/10/2016 | Real Madrid | Osasuna | 5 | 2 | H | 3 | 0 | H | 25 | 10 | 11 | 3 | 14 | 14 | 6 | 7 | 2 | 2 | 0 | 1 | 1.04 | 15 | 34 | 1.06 | 13.5 | 29 |
| 9/10/2016 | Sevilla | Las Palmas | 2 | 1 | H | 0 | 1 | A | 16 | 7 | 7 | 4 | 15 | 13 | 10 | 1 | 2 | 4 | 0 | 1 | 1.5 | 4.33 | 6 | 1.57 | 4.2 | 5.75 |
| 9/11/2016 | Granada | Eibar | 1 | 2 | A | 0 | 1 | A | 15 | 10 | 5 | 8 | 7 | 8 | 6 | 7 | 2 | 0 | 0 | 1 | 2.15 | 3.25 | 3.5 | 2.2 | 3.4 | 3.3 |
| 9/11/2016 | La Coruna | Ath Bilbao | 0 | 1 | A | 0 | 1 | A | 16 | 7 | 3 | 2 | 15 | 15 | 7 | 6 | 2 | 3 | 0 | 3.6 | 3.25 | 2.15 | 3.4 | 3.3 | 2.2 | |
| 9/11/2016 | Sp Gijon | Leganes | 2 | 1 | H | 2 | 0 | H | 7 | 9 | 4 | 2 | 12 | 16 | 3 | 5 | 0 | 2 | 0 | 2.25 | 3.2 | 3.4 | 2.2 | 3.25 | 3.5 | |
| 9/11/2016 | Valencia | Betis | 2 | 3 | A | 0 | 1 | A | 23 | 12 | 9 | 5 | 15 | 21 | 5 | 5 | 2 | 4 | 1 | 1.65 | 4.2 | 5 | 1.7 | 4 | 4.75 | |

FIGURA A.1: Base de datos utilizada. Parte 1.

| B | C | D | ← | → | AF | AG | AH | AI | AJ | AK | AL | AM | AN | AO | AP | AQ | AR | AS | AT | AU | AV | AW | AX | AY |
|-----------|-------------|-------------|---|---|------|------|------|------|-------|-------|------|------|------|------|------|------|-------|-------|-------|-------|-------|-------|-------|------|
| Date | HomeTeam | AwayTeam | | | LBH | LBD | LBA | PSH | PSD | PSA | WHH | WHD | WHA | VCH | VCD | VCA | Bb1X2 | BbMxH | BbAvH | BbMxD | BbAvD | BbMxA | BbAvA | BbOU |
| 8/19/2016 | La Coruna | Eibar | | | 1,95 | 3,25 | 4,2 | 2,01 | 3,39 | 4,39 | 1,95 | 3,25 | 4,2 | 2 | 3,4 | 4,3 | 58 | 2,05 | 1,97 | 3,44 | 3,31 | 4,44 | 4,18 | 43 |
| 8/19/2016 | Malaga | Osasuna | | | 1,6 | 3,6 | 6,5 | 1,61 | 3,89 | 7 | 1,67 | 3,4 | 6 | 1,62 | 3,8 | 6,5 | 57 | 1,7 | 1,62 | 3,89 | 3,66 | 7 | 6,24 | 41 |
| 8/20/2016 | Barcelona | Betis | | | 1,09 | 11 | 29 | 1,1 | 13 | 29,5 | 1,1 | 8 | 26 | 1,07 | 13 | 31 | 57 | 1,15 | 1,09 | 13,25 | 10,4 | 34 | 26,95 | 36 |
| 8/20/2016 | Granada | Villarreal | | | 2,5 | 3 | 2,9 | 2,61 | 3,19 | 3,09 | 2,62 | 3,1 | 2,75 | 2,55 | 3,2 | 3 | 56 | 2,62 | 2,54 | 3,21 | 3,13 | 3,11 | 2,92 | 40 |
| 8/20/2016 | Sevilla | Espanol | | | 1,67 | 3,8 | 4,75 | 1,73 | 4,02 | 5,18 | 1,73 | 3,5 | 5 | 1,73 | 4 | 5 | 57 | 1,73 | 1,68 | 4,07 | 3,86 | 5,5 | 5,04 | 44 |
| 8/21/2016 | Ath Madrid | Alaves | | | 1,22 | 5,5 | 17 | 1,23 | 6,25 | 20,4 | 1,2 | 6 | 15 | 1,22 | 6,25 | 19 | 56 | 1,28 | 1,21 | 6,5 | 5,88 | 21 | 16,97 | 43 |
| 8/21/2016 | Sociedad | Real Madrid | | | 4,5 | 3,8 | 1,75 | 5,13 | 3,98 | 1,74 | 4,5 | 3,5 | 1,8 | 4,8 | 3,9 | 1,75 | 57 | 5,13 | 4,66 | 4 | 3,81 | 1,8 | 1,74 | 40 |
| 8/21/2016 | Sp Gijon | Ath Bilbao | | | 4 | 3,4 | 1,91 | 4,4 | 3,54 | 1,96 | 3,75 | 3,4 | 2 | 4,3 | 3,5 | 1,95 | 56 | 4,4 | 4,06 | 3,6 | 3,41 | 2 | 1,94 | 42 |
| 8/22/2016 | Celta | Leganes | | | 1,6 | 3,75 | 6 | 1,63 | 4 | 6,47 | 1,55 | 4 | 6 | 1,62 | 4 | 6,25 | 54 | 1,67 | 1,6 | 4 | 3,83 | 6,47 | 5,89 | 41 |
| 8/22/2016 | Valencia | Las Palmas | | | 1,73 | 3,6 | 4,75 | 1,77 | 3,89 | 5,07 | 1,65 | 4 | 5 | 1,75 | 3,8 | 5 | 55 | 1,79 | 1,73 | 4 | 3,71 | 5,07 | 4,84 | 42 |
| 8/26/2016 | Betis | La Coruna | | | 2,15 | 3,2 | 3,5 | 2,22 | 3,34 | 3,7 | 2,25 | 3,2 | 3,3 | 2,25 | 3,3 | 3,5 | 54 | 2,33 | 2,21 | 3,37 | 3,22 | 3,7 | 3,46 | 43 |
| 8/26/2016 | Espanol | Malaga | | | 2,3 | 3,1 | 3,25 | 2,38 | 3,27 | 3,39 | 2,3 | 3,2 | 3,25 | 2,3 | 3,25 | 3,4 | 54 | 2,4 | 2,32 | 3,3 | 3,18 | 3,41 | 3,26 | 42 |
| 8/27/2016 | Eibar | Valencia | | | 2,62 | 3,1 | 2,7 | 2,78 | 3,25 | 2,84 | 2,62 | 3,1 | 2,8 | 2,75 | 3,25 | 2,8 | 54 | 2,8 | 2,67 | 3,29 | 3,18 | 2,85 | 2,75 | 42 |
| 8/27/2016 | Leganes | Ath Madrid | | | 7 | 3,6 | 1,55 | 7,88 | 3,83 | 1,58 | 7,5 | 3,5 | 1,57 | 8 | 3,8 | 1,57 | 53 | 8,5 | 7,32 | 3,9 | 3,65 | 1,62 | 1,56 | 38 |
| 8/27/2016 | Osasuna | Sociedad | | | 3,1 | 3 | 2,5 | 3,19 | 3,13 | 2,59 | 3 | 3 | 2,5 | 3,13 | 3,13 | 2,55 | 54 | 3,19 | 3,04 | 3,2 | 3,06 | 2,65 | 2,52 | 39 |
| 8/27/2016 | Real Madrid | Celta | | | 1,2 | 7 | 12 | 1,22 | 7,9 | 14,09 | 1,2 | 7 | 13 | 1,2 | 7,5 | 15 | 54 | 1,27 | 1,21 | 7,9 | 6,98 | 17 | 12,8 | 34 |
| 8/28/2016 | Alaves | Sp Gijon | | | 2,1 | 3,1 | 3,8 | 2,15 | 3,26 | 4,02 | 2,15 | 3,2 | 3,5 | 2,1 | 3,25 | 3,75 | 52 | 2,2 | 2,14 | 3,26 | 3,2 | 4,02 | 3,68 | 37 |
| 8/28/2016 | Ath Bilbao | Barcelona | | | 5,75 | 4,33 | 1,53 | 6,44 | 4,66 | 1,54 | 6 | 4,33 | 1,5 | 6 | 4,6 | 1,5 | 53 | 6,7 | 6,09 | 4,9 | 4,41 | 1,6 | 1,52 | 38 |
| 8/28/2016 | Las Palmas | Granada | | | 1,85 | 3,4 | 4,33 | 1,91 | 3,6 | 4,58 | 1,95 | 3,4 | 4 | 1,87 | 3,6 | 4,2 | 53 | 2,02 | 1,9 | 3,6 | 3,45 | 4,58 | 4,25 | 42 |
| 8/28/2016 | Villarreal | Sevilla | | | 2,5 | 3,3 | 2,8 | 2,6 | 3,42 | 2,91 | 2,5 | 3,3 | 2,8 | 2,45 | 3,4 | 2,88 | 53 | 2,6 | 2,5 | 3,5 | 3,33 | 2,92 | 2,84 | 42 |
| 9/9/2016 | Sociedad | Espanol | | | 2 | 3,3 | 3,8 | 2,06 | 3,53 | 4 | 2,05 | 3,4 | 3,6 | 2,05 | 3,5 | 3,9 | 56 | 2,1 | 2,03 | 3,56 | 3,41 | 4,01 | 3,82 | 44 |
| 9/10/2016 | Barcelona | Alaves | | | 1,08 | 12 | 34 | 1,1 | 12,25 | 34 | 1,08 | 11 | 26 | 1,06 | 13 | 36 | 56 | 1,14 | 1,07 | 14,5 | 11,58 | 46 | 32,36 | 37 |
| 9/10/2016 | Celta | Ath Madrid | | | 5,25 | 3,3 | 1,75 | 5,64 | 3,59 | 1,78 | 4,8 | 3,4 | 1,8 | 5,25 | 3,5 | 1,8 | 53 | 6 | 5,25 | 3,6 | 3,43 | 1,81 | 1,77 | 40 |
| 9/10/2016 | Malaga | Villarreal | | | 2,3 | 3,1 | 3,3 | 2,39 | 3,24 | 3,41 | 2,5 | 3,2 | 2,88 | 2,38 | 3,2 | 3,4 | 56 | 2,5 | 2,37 | 3,3 | 3,15 | 3,42 | 3,21 | 40 |
| 9/10/2016 | Real Madrid | Osasuna | | | 1,06 | 15 | 41 | 1,07 | 19 | 33,5 | 1,07 | 11 | 29 | 1,05 | 19 | 46 | 56 | 1,13 | 1,06 | 19 | 13,74 | 55 | 36,19 | 38 |
| 9/10/2016 | Sevilla | Las Palmas | | | 1,53 | 4,33 | 5,75 | 1,55 | 4,66 | 6,27 | 1,57 | 3,6 | 6,5 | 1,6 | 4,6 | 5,4 | 56 | 1,7 | 1,56 | 4,73 | 4,31 | 6,5 | 5,68 | 38 |
| 9/11/2016 | Granada | Eibar | | | 2,15 | 3,3 | 3,3 | 2,23 | 3,49 | 3,5 | 2,25 | 3,2 | 3,3 | 2,2 | 3,4 | 3,3 | 56 | 2,34 | 2,22 | 3,5 | 3,36 | 3,53 | 3,3 | 45 |
| 9/11/2016 | La Coruna | Ath Bilbao | | | 3,4 | 3,2 | 2,2 | 3,55 | 3,38 | 2,26 | 3,3 | 3,2 | 2,25 | 3,5 | 3,25 | 2,2 | 55 | 3,66 | 3,43 | 3,45 | 3,28 | 2,26 | 2,2 | 42 |
| 9/11/2016 | Sp Gijon | Leganes | | | 2,2 | 3,1 | 3,4 | 2,29 | 3,31 | 3,56 | 2,3 | 3,2 | 3,2 | 2,25 | 3,3 | 3,5 | 55 | 2,32 | 2,24 | 3,31 | 3,22 | 3,62 | 3,42 | 40 |
| 9/11/2016 | Valencia | Betis | | | 1,65 | 3,8 | 5,25 | 1,68 | 4,08 | 5,53 | 1,73 | 3,5 | 5 | 1,67 | 4,1 | 5,4 | 56 | 1,75 | 1,68 | 4,2 | 3,88 | 5,53 | 5,05 | 46 |

FIGURA A.2: Base de datos utilizada. Parte 2.

| B | C | D | AZ | BA | BB | BC | BD | BE | BF | BG | BH | BI | BJ | BK | BL |
|-----------|-------------|-------------|----------|----------|----------|----------|------|-------|---------|---------|---------|---------|------|-------|-------|
| Date | HomeTeam | AwayTeam | BbMx>2,5 | BbAv>2,5 | BbMx<2,5 | BbAv<2,5 | BbAH | BbAHh | BbMxAHH | BbAvAHH | BbMxAHA | BbAvAHA | PSCH | PSCD | PSCA |
| 8/19/2016 | La Coruna | Eibar | 2,47 | 2,34 | 1,63 | 1,58 | 33 | -0,5 | 2,01 | 1,95 | 1,94 | 1,89 | 2,05 | 3,39 | 4,2 |
| 8/19/2016 | Malaga | Osasuna | 2,44 | 2,31 | 1,65 | 1,6 | 34 | -1 | 2,29 | 2,11 | 1,82 | 1,76 | 1,66 | 3,78 | 6,59 |
| 8/20/2016 | Barcelona | Betis | 1,31 | 1,29 | 3,8 | 3,52 | 34 | -2,5 | 1,87 | 1,78 | 2,15 | 2,08 | 1,1 | 12,65 | 29,5 |
| 8/20/2016 | Granada | Villarreal | 2,42 | 2,33 | 1,65 | 1,58 | 30 | 0 | 1,83 | 1,79 | 2,15 | 2,07 | 2,76 | 3,22 | 2,89 |
| 8/20/2016 | Sevilla | Espanol | 1,82 | 1,74 | 2,17 | 2,07 | 36 | -1 | 2,26 | 2,16 | 1,84 | 1,72 | 1,71 | 3,93 | 5,49 |
| 8/21/2016 | Ath Madrid | Alaves | 2,02 | 1,94 | 1,92 | 1,85 | 34 | -1,5 | 1,82 | 1,76 | 2,17 | 2,09 | 1,24 | 6 | 20 |
| 8/21/2016 | Sociedad | Real Madrid | 1,83 | 1,73 | 2,25 | 2,09 | 33 | 1 | 1,68 | 1,64 | 2,38 | 2,29 | 5,49 | 4,02 | 1,69 |
| 8/21/2016 | Sp Gijon | Ath Bilbao | 2,24 | 2,15 | 1,76 | 1,69 | 32 | 0,5 | 1,98 | 1,91 | 1,98 | 1,94 | 4,66 | 3,57 | 1,9 |
| 8/22/2016 | Celta | Leganes | 2,13 | 2,02 | 1,88 | 1,77 | 34 | -1 | 2,16 | 2,07 | 1,9 | 1,78 | 1,68 | 3,84 | 6,09 |
| 8/22/2016 | Valencia | Las Palmas | 1,98 | 1,9 | 1,96 | 1,88 | 33 | -1 | 2,45 | 2,32 | 1,68 | 1,62 | 1,75 | 3,89 | 5,23 |
| 8/26/2016 | Betis | La Coruna | 2,34 | 2,21 | 1,7 | 1,64 | 30 | -0,25 | 1,92 | 1,88 | 2,05 | 1,98 | 2,21 | 3,37 | 3,69 |
| 8/26/2016 | Espanol | Malaga | 2,42 | 2,32 | 1,65 | 1,59 | 30 | -0,25 | 2,03 | 1,98 | 1,92 | 1,88 | 2,34 | 3,25 | 3,51 |
| 8/27/2016 | Eibar | Valencia | 2,3 | 2,14 | 1,75 | 1,69 | 29 | 0,25 | 1,67 | 1,63 | 2,4 | 2,32 | 2,84 | 3,34 | 2,72 |
| 8/27/2016 | Leganes | Ath Madrid | 2,6 | 2,49 | 1,57 | 1,51 | 33 | 1 | 1,88 | 1,83 | 2,1 | 2,03 | 9,31 | 4,02 | 1,51 |
| 8/27/2016 | Osasuna | Sociedad | 2,7 | 2,52 | 1,57 | 1,5 | 28 | 0,25 | 1,8 | 1,75 | 2,23 | 2,13 | 3,12 | 3,22 | 2,57 |
| 8/27/2016 | Real Madrid | Celta | 1,37 | 1,32 | 3,5 | 3,27 | 34 | -2 | 1,99 | 1,92 | 1,98 | 1,93 | 1,18 | 8,5 | 17,85 |
| 8/28/2016 | Alaves | Sp Gijon | 2,6 | 2,44 | 1,57 | 1,53 | 29 | -0,25 | 1,9 | 1,84 | 2,12 | 2,02 | 2,52 | 3,14 | 3,28 |
| 8/28/2016 | Ath Bilbao | Barcelona | 1,64 | 1,56 | 2,55 | 2,35 | 33 | 1 | 2,06 | 2 | 1,9 | 1,85 | 6,33 | 4,37 | 1,58 |
| 8/28/2016 | Las Palmas | Granada | 2,11 | 2,03 | 1,83 | 1,76 | 28 | -0,25 | 1,68 | 1,63 | 2,42 | 2,32 | 1,83 | 3,85 | 4,7 |
| 8/28/2016 | Villarreal | Sevilla | 1,9 | 1,83 | 2,05 | 1,94 | 28 | 0 | 1,86 | 1,8 | 2,13 | 2,05 | 2,4 | 3,5 | 3,15 |
| 9/9/2016 | Sociedad | Espanol | 2,15 | 2,04 | 1,84 | 1,76 | 31 | -0,75 | 2,43 | 2,35 | 1,67 | 1,62 | 1,99 | 3,62 | 4,14 |
| 9/10/2016 | Barcelona | Alaves | 1,3 | 1,26 | 4,1 | 3,74 | 29 | -2,5 | 1,7 | 1,66 | 2,35 | 2,25 | 1,09 | 13,38 | 32,32 |
| 9/10/2016 | Celta | Ath Madrid | 2,66 | 2,5 | 1,59 | 1,51 | 31 | 0,75 | 1,88 | 1,84 | 2,08 | 2,02 | 6,1 | 3,61 | 1,73 |
| 9/10/2016 | Malaga | Villarreal | 2,61 | 2,46 | 1,58 | 1,53 | 30 | -0,25 | 2,08 | 2 | 1,91 | 1,86 | 2,53 | 3,11 | 3,3 |
| 9/10/2016 | Real Madrid | Osasuna | 1,24 | 1,2 | 4,7 | 4,3 | 33 | -3 | 1,85 | 1,8 | 2,13 | 2,06 | 1,07 | 1,77 | 36,4 |
| 9/10/2016 | Sevilla | Las Palmas | 1,55 | 1,5 | 2,66 | 2,54 | 36 | -1 | 1,95 | 1,87 | 2,05 | 1,98 | 1,53 | 4,61 | 6,71 |
| 9/11/2016 | Granada | Eibar | 2,06 | 1,98 | 1,9 | 1,81 | 32 | -0,25 | 2,02 | 1,91 | 2,01 | 1,95 | 2,58 | 3,24 | 3,09 |
| 9/11/2016 | La Coruna | Ath Bilbao | 2,45 | 2,32 | 1,64 | 1,59 | 32 | 0,25 | 2,02 | 1,97 | 1,93 | 1,88 | 3,68 | 3,3 | 2,25 |
| 9/11/2016 | Sp Gijon | Leganes | 2,54 | 2,41 | 1,6 | 1,55 | 32 | -0,25 | 1,95 | 1,91 | 2 | 1,94 | 2,63 | 3,07 | 3,18 |
| 9/11/2016 | Valencia | Betis | 1,9 | 1,83 | 2,05 | 1,96 | 34 | -1 | 2,26 | 2,18 | 1,75 | 1,71 | 1,55 | 4,42 | 6,74 |

FIGURA A.3: Base de datos utilizada. Parte 3.

Apéndice B

Planificación

B.1. Calendario teórico

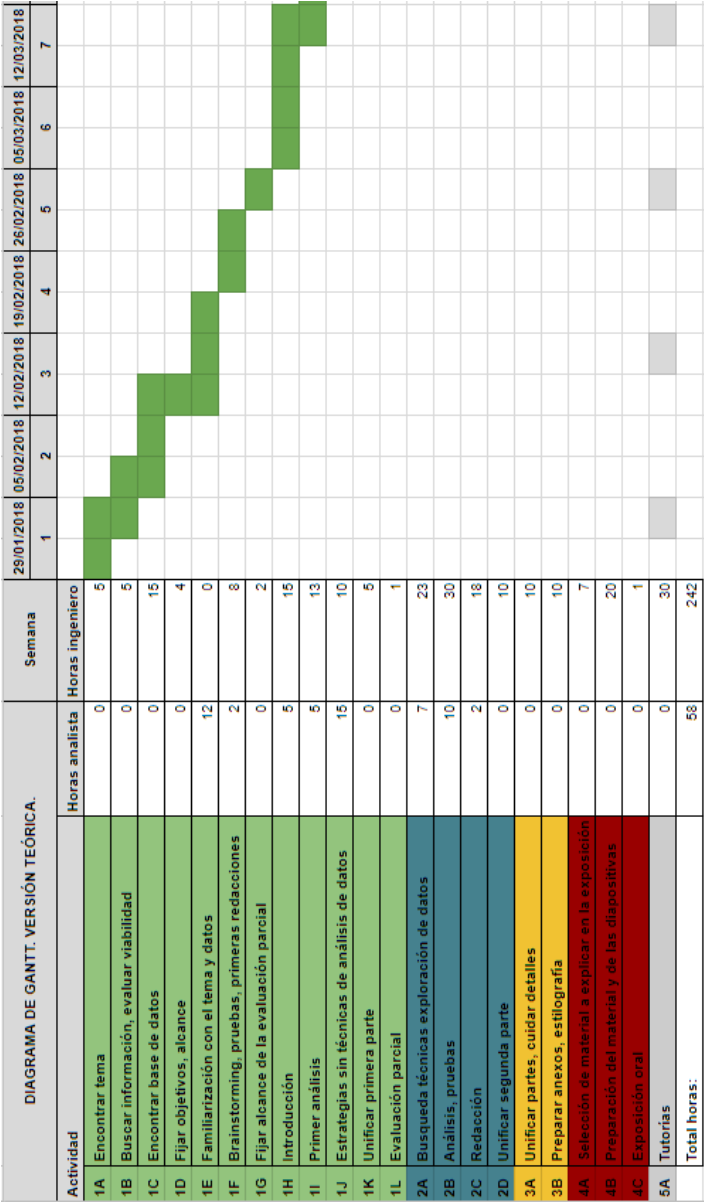


FIGURA B.1: Diagrama de Gantt teórico. Parte 1.

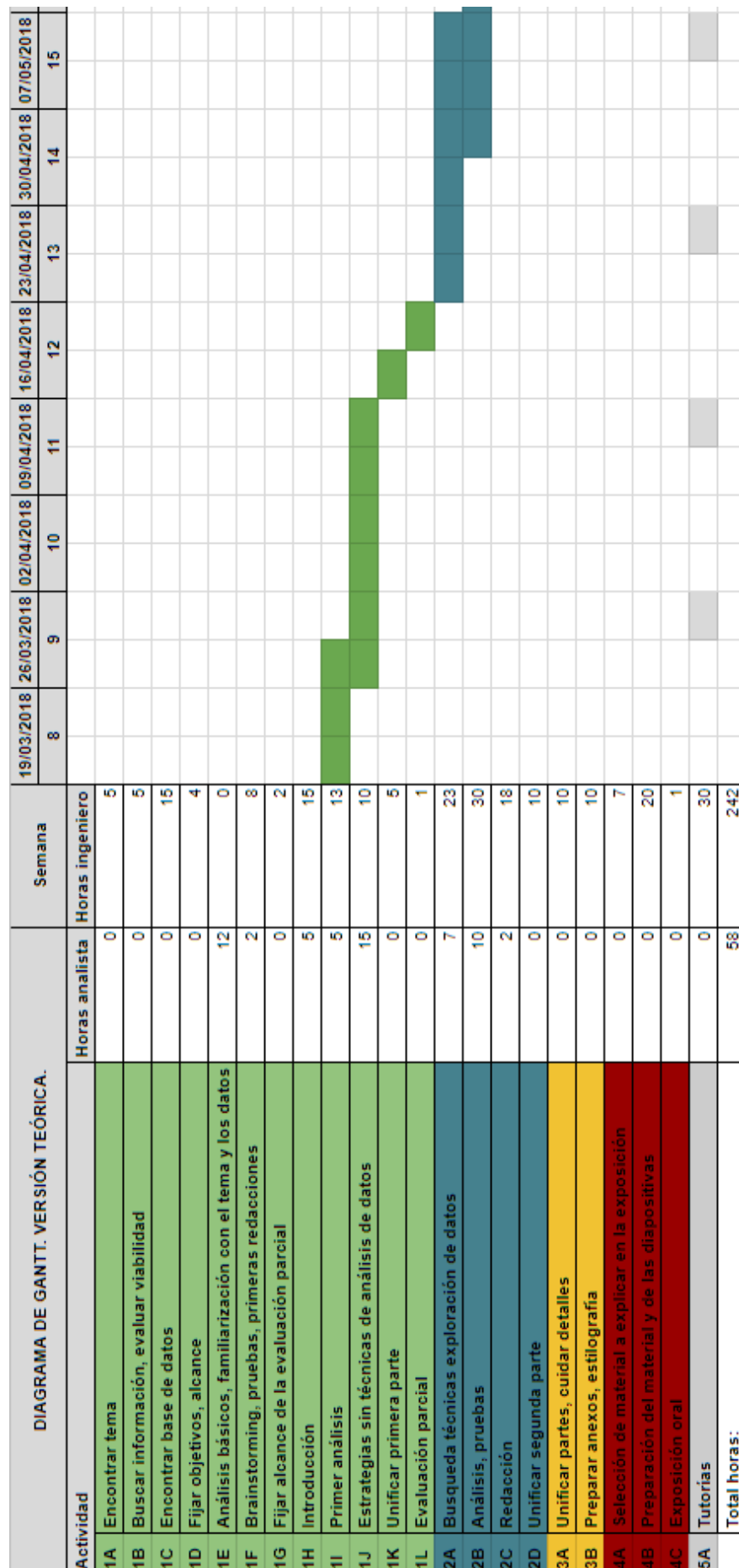


FIGURA B.2: Diagrama de Gantt teórico. Parte 2.

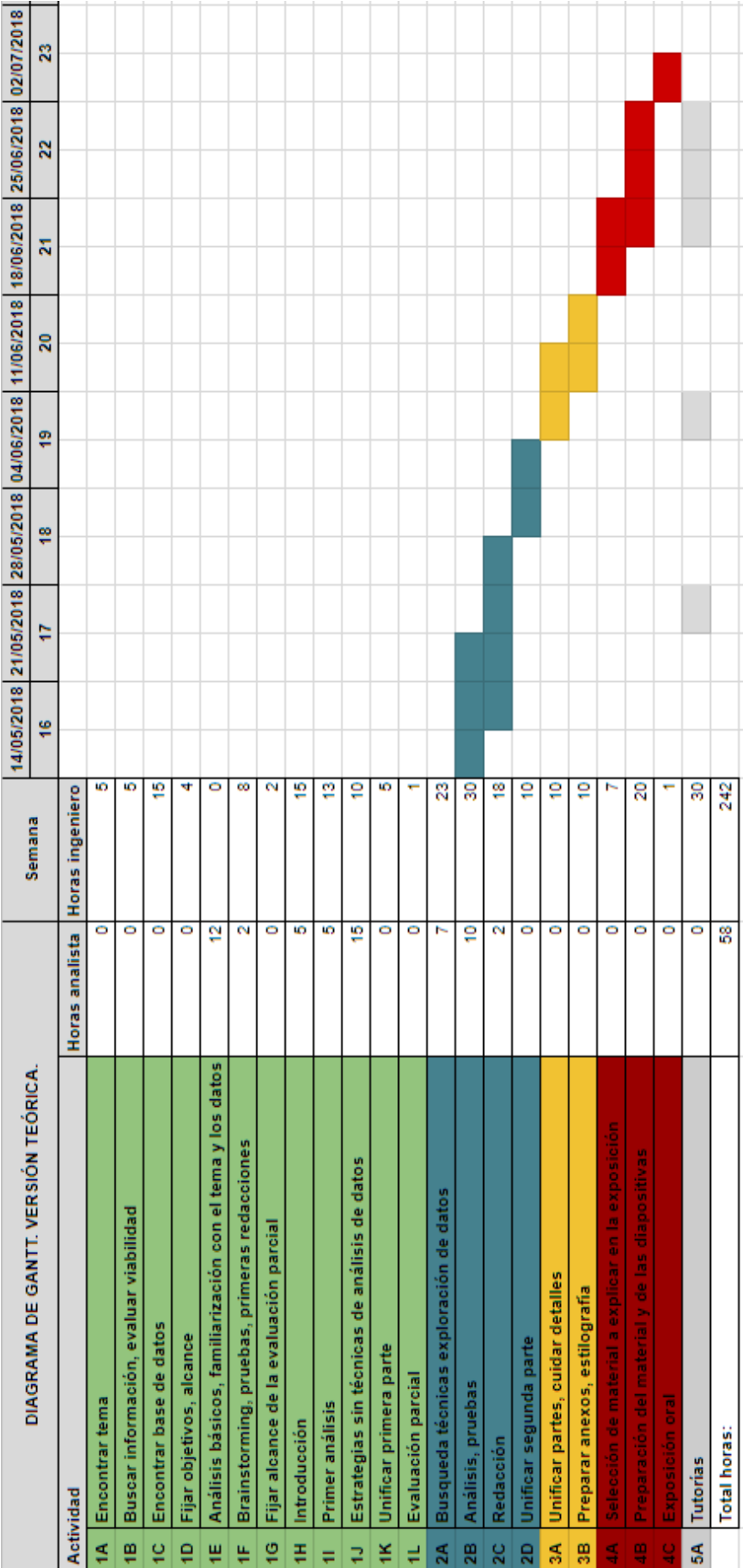


FIGURA B.3: Diagrama de Gantt teórico. Parte 3.

B.2. Calendario práctico

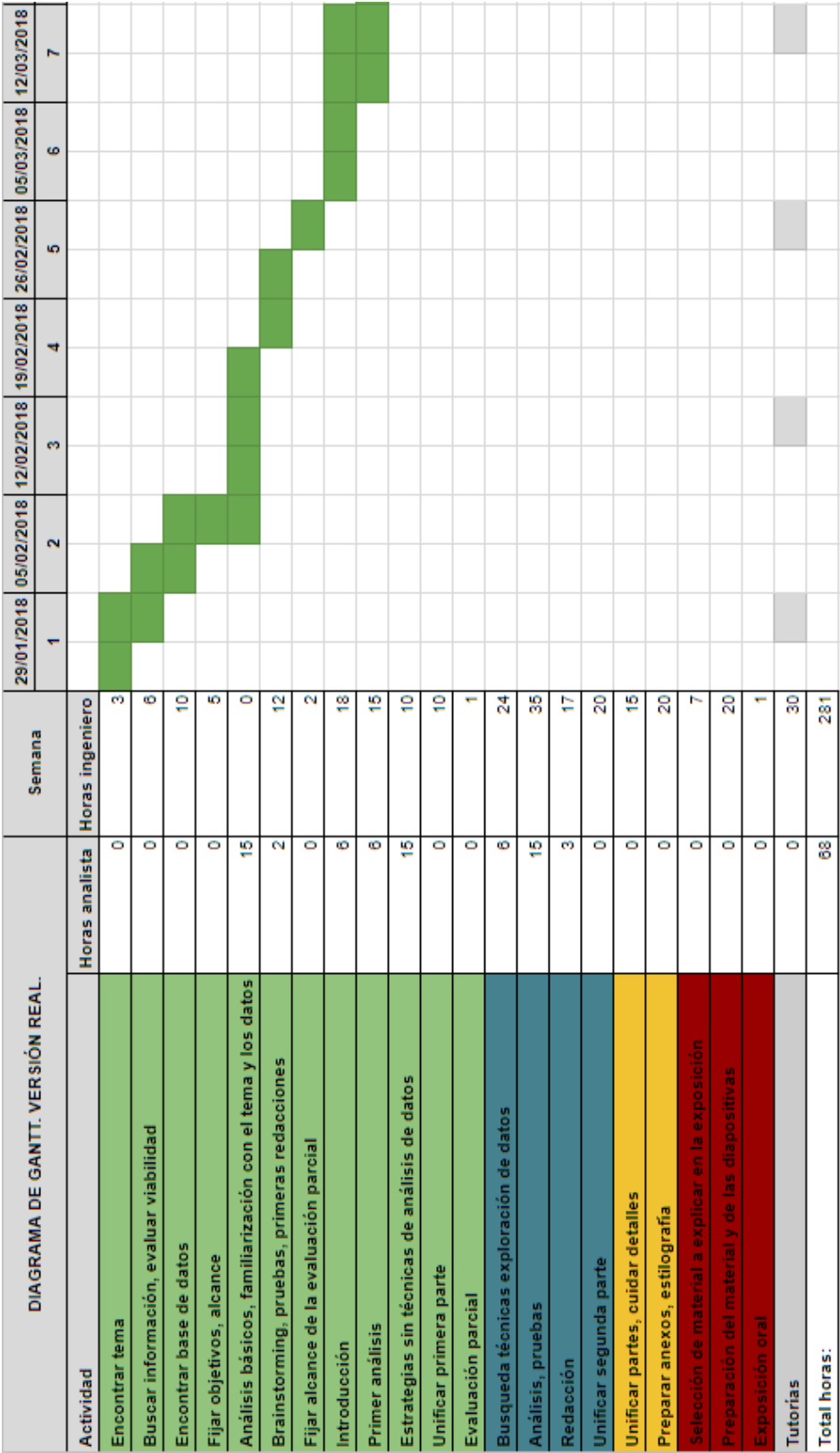


FIGURA B.4: Diagrama de Gantt práctico. Parte 1.

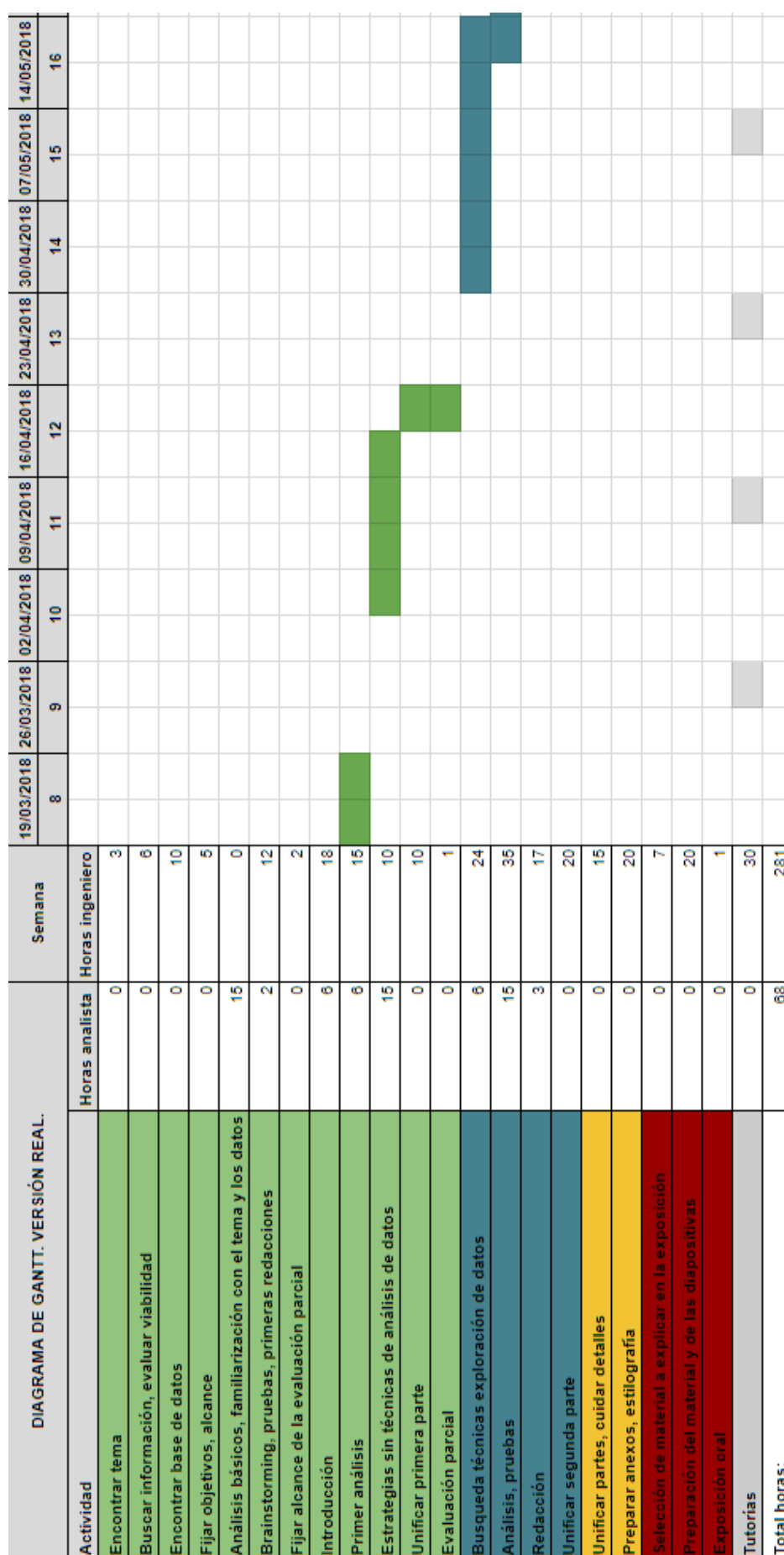


FIGURA B.5: Diagrama de Gantt práctico. Parte 2.

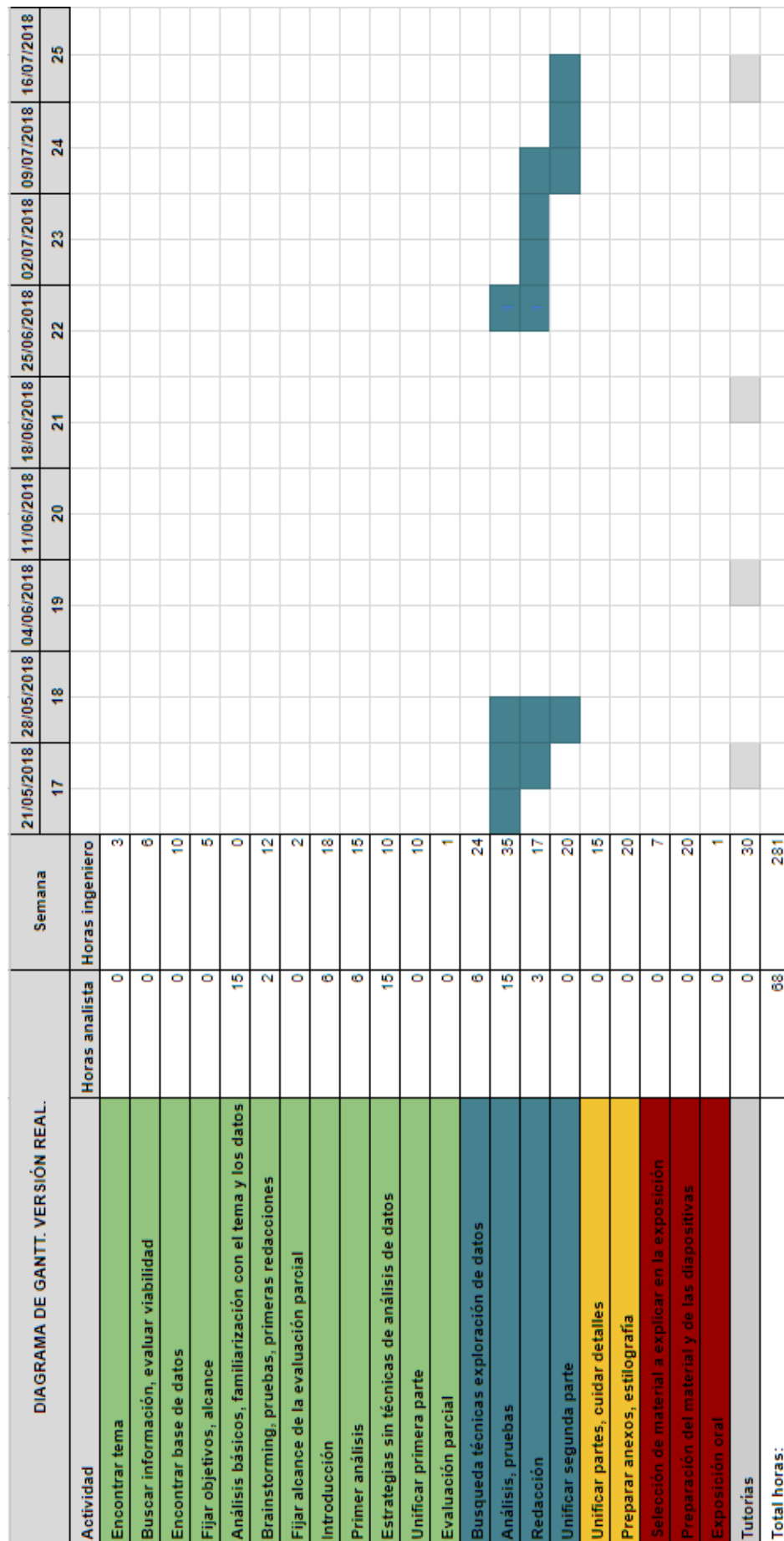


FIGURA B.6: Diagrama de Gantt práctico. Parte 3.

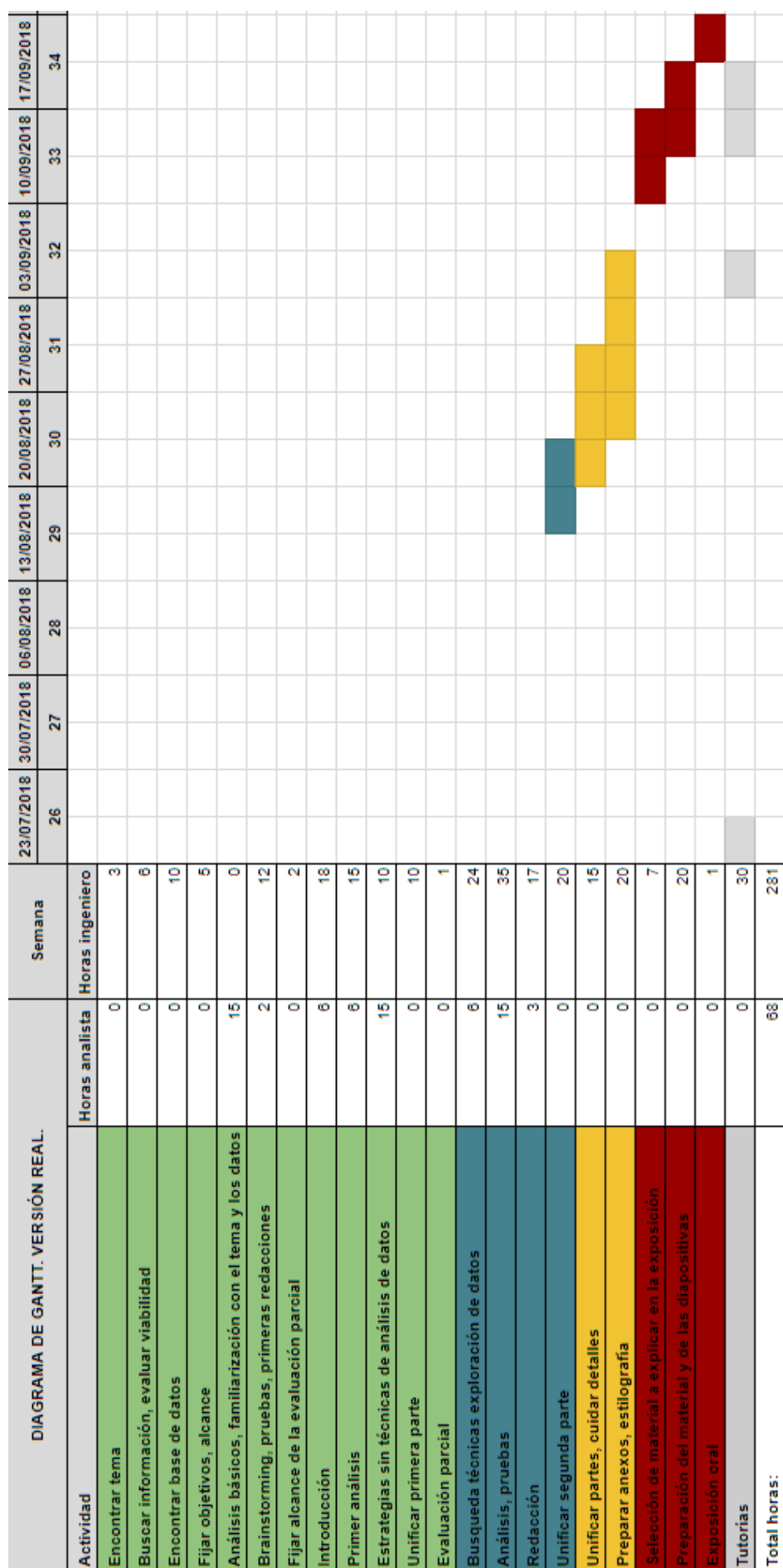


FIGURA B.7: Diagrama de Gantt práctico. Parte 4.